IRE Transactions

ON AUTOMATIC CONTROL

D. R. Ziemer

PGAC-1

MAY, 1956

TABLE OF CONTENTS

PUBLISHED BY THE

PROFESSIONAL GROUP ON AUTOMATIC CONTROL

IRE PROFESSIONAL GROUP ON AUTOMATIC CONTROL

The Professional Group on Automatic Control is an organization, within the framework of the IRE, of members with principal professional interest in Automatic Control. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of prescribed assessments.

Annual Assessment: $2.00

IRE TRANSACTIONS®
on Automatic Control

Published by the Institute of Radio Engineers, Inc., for the Professional Group on Automatic Control, 1 East 79th Street, New York 21, New York. Responsibility for the contents rests upon the authors, and not upon the IRE, the Group or its members. Individual copies available for sale to IRE-PGAC members at $1.95, to IRE members at $2.90 and to nonmembers at $5.85.

AUTOMATIC CONTROL AND THE IRE

This is the first Transactions of the Professional Group on Automatic Control. Like the Group it is small, and its infancy is reflected within by the typewritten, offset printing. When the Group grows larger, more mature and increasingly active, the Transactions will be typeset and more attractive.

Although the PGAC is young, the IRE actually has been active in the field of automatic control for a number of years. In the past some of the first theories of feedback control were developed by those interested in relatively simple electronic feedback amplifiers. Radio and electronics engineers were among those who created new theories and developed more advanced feedback circuits until today radios, radar, television and computers contain numerous complex feedback control circuits which include semiconductors, electromagnetic devices and mechanical elements as well as vacuum tubes.

And a new trend is developing -- all of the complex marvels, the radio, the radar, the television, the computer, each containing internal feedback control loops, are being combined into over-all automatic control systems. These have been referred to by the military in the weapon systems concept and by industry as automation. Unique problems have appeared during the integration of these complex systems; new theories such as those concerning the nature of information which must flow between system elements, and new techniques such as those involved in digital computation are being developed.

In the future automatic control will become more complex; it will encompass broader fields. There will be a greater need for developing and incorporating new theories, techniques and components into integrated systems. The IRE may well participate in the evolution of these integrated systems, not only because of the past association with feedback control but because at the present time many of the new techniques and theories necessary for the future development of automatic control are being fostered by other professional groups in the IRE. It will be a function of the future automatic control engineer to pose new problems for specialists in the related groups and to coordinate new information into integrated, automatically controlled systems.

To represent the automatic control engineer, to provide a means for disseminating information about automatic control, its related interests, its problems and its developments, the Professional Group on Automatic Control was formed. From the constitution of the PGAC: "...The field of interest of the Group shall be automatic control systems. It shall encompass the components thereof, such as transducers, data transmission links, computers and control devices and the integration of these components into control systems...."

Although the interest of the Group will concern the problems of combining elements into over-all complex systems and the solutions of these problems, attention will be given to any subject, abstract or practical, relating to automatic control within any of the system elements, simple or complex. It should be observed that the field of interest is broad and the subjects to be considered varied. This first Transactions issue illustrates the range of interest. Although there are only seven papers, the topics covered are diversified, from the first which is a specific example of the integration of a digital computer into a feedback control system, to the last which is a general discussion of the future of automatic control and its relationship to the engineer.

The intervening papers concern linear and nonlinear theory applied to simple feedback systems and to practical solutions of problems inherent in the development of a complex fire control system.

Each of the papers was selected from a large number by a Papers Study and Procurement Committee. Although most of these papers were conference papers, not specifically intended for the Transactions, others are being obtained from local PGAC chapters where they may or may not have been formally presented. Actually, any papers concerning automatic control will be welcome, but they will have to be reviewed by the committee before being published in the Transactions. Rejected manuscripts will be returned to the author with comments and suggested revisions. To aid in preparing papers and to minimize rejects, a suggested form for writing manuscripts is included in this issue. Papers should be submitted to Mr. M. R. Aaron, Bell Telephone Laboratories, Murray Hill, New Jersey.

The growth of the PGAC and the aid which the IRE can give to the development of automatic control will depend considerably on the activity of local PGAC chapters and the dissemination of information within the Group and among other groups. The Transactions is the medium whereby this information can be distributed; it is the agent where new problems and developments can be proposed. Its usefulness will depend on the material it will contain: articles on general theory and new practical applications, tutorial articles, historical surveys, design methods, discussions of papers already presented, abstracts of papers printed in other group Transactions which may be of interest to PGAC members, book reviews, chapter activities or theoretical problems for educational purposes. Comments and suggestions for improving the Transactions in any way, by changing the format or in the selection of material, will be most welcome -- and necessary -- if the Transactions is to serve its purpose in the development of automatic control.

SYNTHESIS OF A NONLINEAR CONTROL SYSTEM*†

I. Flügge-Lotz and C. F. Taylor
Stanford University
Stanford, California

Summary -- The investigated nonlinear control system consists of a linear member with an ensemble of possible discrete combinations of proportional and derivative feedback around the linear member. The particular combination of proportional and derivative feedback employed at any instant is determined by a feedback switching circuit which is in turn operated by sensed binary information obtained from the output, output derivative, error and error derivative [namely, the signs (sgn) of these variables]. Techniques that are common to the digital computer field are used to implement this switching circuit.

For a linear member of second order the feedback circuit is comprised of four discrete values of proportional feedback and four discrete values of derivative feedback.

Simulation techniques have been used to study and evaluate the performance of the nonlinear control system and to compare it with a linear system for a wide variety of inputs. A sample of these experimental results is presented. The experiments show that the nonlinear system performance is much better than that of a linear system of comparable power handling capability.

Preliminary studies of a nonlinear control system of third order show that the basic idea can be extended to higher order systems.

The decision to introduce nonlinear components in a control system may stem from the desire to have a system whose performance is less tied to history than that of linear control systems of comparable power handling capability.

Such a nonlinear system has been mathematically described earlier.‡ In the present paper its physical realization and experiments for studying performance are described.

The particular control process may be visualized as

(1) establishing an ensemble of linear differential equations with constant coefficients (some with stable and some with unstable homogeneous solutions) and

(2) switching from one member of the ensemble to another so as to maintain small instantaneous error between input and output for relatively arbitrary inputs. The switching sequence and time of switching are functionally dependent upon quantized binary information of the output

*Research supported by National Advisory Committee for Aeronautics (NACA).
†This paper was presented at WESCON, the West Coast Conference sponsored by the IRE - West Coast Electronics Manufacturer's Association, in August, 1955.
‡Flügge-Lotz, I., and Wunch, W. S.: "On a nonlinear transfer system." Jour. Appl. Phys., pp. 484-488; Apr., 1955.

and input, namely the signs (sgn) of the output, output derivative, error and error derivative.

Briefly, the physical manifestation of this control process is a linear system with discrete variable proportional and derivative feedback around the linear member. The particular combination of proportional and derivative feedback employed at any instant is determined by a binary logic switching circuit. This circuit operates with sensed quantized information of the output and input.

MATHEMATICAL DESCRIPTION OF A SECOND ORDER NONLINEAR CONTROL SYSTEM

Mathematically, the system to be considered is described by

$$\frac{D^2y}{dt^2} + 2D(1 + \beta_m)\frac{dy}{dt} + (1 + \gamma_n)y = x(t) \tag{1}$$

where $x(t)$ is the input, $y(t)$ is the output and D is the linear damping factor (when $\beta_m = \gamma_n = 0$). The instantaneous error is defined as $e = x - y$. The parameters β_m and γ_n are stepwise switching functions of their implicit variable, time, and are given explicitly as functions of x and y. That is,

$$\beta_m = -\mathrm{sgn}(y') \left[{}_1\beta \,\mathrm{sgn}(e) + {}_2\beta \,\mathrm{sgn}(e')\right]$$

$$\left.\begin{aligned} \beta_3 &= {}_1\beta + {}_2\beta \\ \beta_2 &= {}_1\beta - {}_2\beta \\ \beta_1 &= -{}_1\beta + {}_2\beta = -\beta_2 \\ \beta_0 &= -{}_1\beta - {}_2\beta = -\beta_3 \end{aligned}\right\} \begin{array}{l} {}_1\beta \text{ and } {}_2\beta \text{ are} \\ \text{positive constants} \end{array}$$

$$\gamma_n = -\mathrm{sgn}(y) \left[{}_1\gamma \,\mathrm{sgn}(e) + {}_2\gamma \,\mathrm{sgn}(e')\right]$$

$$\left.\begin{aligned} \gamma_3 &= {}_1\gamma + {}_2\gamma \\ \gamma_2 &= {}_1\gamma - {}_2\gamma \\ \gamma_1 &= -{}_1\gamma + {}_2\gamma = -\gamma_2 \\ \gamma_0 &= -{}_1\gamma - {}_2\gamma = -\gamma_3 \end{aligned}\right\} \begin{array}{l} {}_1\gamma \text{ and } {}_2\gamma \text{ are} \\ \text{positive constants} \end{array}$$

and

$$\mathrm{sgn}(f) \triangleq \frac{f}{|f|} = \begin{cases} +1 \text{ for } f > 0 \\ -1 \text{ for } f < 0 \end{cases}$$

PHYSICAL REALIZATION

In order to understand the nature of a physical system described by Eq. (1), this equation is rearranged and rewritten in operational form (utilizing the Heaviside operator $p = d/dt$) as

$$\left[p^2 + 2Dp + 1\right] y \overset{\Omega}{=} x - \left[2D\beta_m p + \gamma_n\right] y \tag{2}$$

From Eq. (2) a symbolic representation of the nonlinear system is obtained in the block diagram form as shown in Fig. 1.

With the aid of the block diagram the interpretation of Eq. (1) is straightforward. The nonlinear system consists of

(1) a linear second order member. This could be a simple position servo, for example.

(2) a feedback circuit comprised of

  (a) four discrete values of proportional feedback (two positive and two negative) denoted by gain constants, γ_n.

  (b) four discrete values of derivative feedback (two positive and two negative) denoted by gain constants, β_m.

  The particular (β_m, γ_n) feedback combination employed at any instant is determined by a switching logic circuit.

SWITCHING LOGIC

In the physical realization of Eq. (1) a switching logic has to be established that ensures that the proper (β_m, γ_n) combination is employed at the proper time. By definition, the parameters β_m and γ_n are functionally dependent upon the signs (sgn) of the four variables y, y', e and e'. Since the sign of a variable is binary information it is convenient to utilize digital

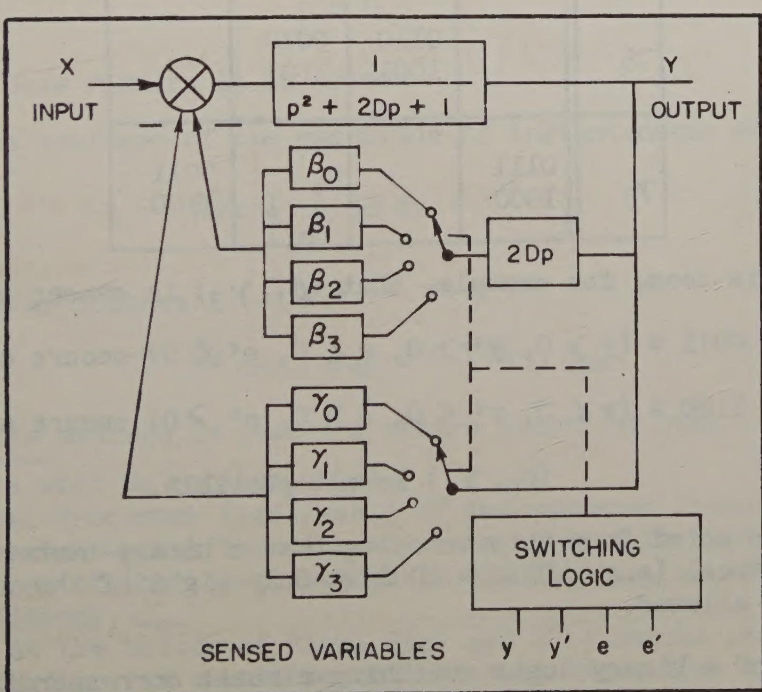


Fig. 1 - Block diagram of Eq. (1).

computer techniques in establishing the necessary switching logic. This is done as follows:

Let $y > 0$ be represented by "0" (binary zero), and

$y < 0$ be represented by "1" (binary one).

Similarly for y', e and e'.

If the ordered sequence is established as

$$[y, y', e, e']$$

then the allowed (β_m, γ_n) combinations may be encoded in four digit binary logic (namely, binary coded decimal). This is done in Table I.

TABLE I
Matrix of Allowed (β_m, γ_n) Combinations Encoded in Binary Logic

	β_0	β_1	β_2	β_3
γ_0	0000 1111			0100 1011
γ_1		0001 1110	0101 1010	
γ_2		0110 1001	0010 1101	
γ_3	0111 1000			0011 1100

From Table I it is seen, for example, that (β_3, γ_3) is chosen when

$0011 = (y > 0,\ y' > 0,\ e < 0,\ e' < 0)$ occurs or

$1100 = (y < 0,\ y' < 0,\ e > 0,\ e' > 0)$ occurs and that

(β_0, γ_1) is not possible.

In general, it is noted from this encoding that a binary number and its complement are identical (e.g., 0110 = 1001). Only eight of the sixteen (β_m, γ_n) conbinations are allowed.

The design of a binary logic switching circuit corresponding to Table I is straightforward. The signs of the four variables y, y', e and e' are sensed and on the basis of the 2^4 binary decisions the allowed (β_m, γ_n) com-

binations are connected as required. Such a switching circuit has been constructed and used in conjunction with an analog computer for simulation studies of this type of nonlinear control system. The only essential components required are relays for switching and zero-coincidence detectors (amplitude selectors) for sign sensing (i.e., to "read in" the binary logic of Table I)

EXPERIMENTAL STUDIES

In the original analytical derivation of Eq. (1) it was not possible to determine analytically the actual magnitudes of the β_m and γ_n parameters for optimum performance. Further, it was assumed that the switching from one parameter set to another would be done ideally (i.e., instantaneously). However, once Eq. (1) has been synthesized, simulation techniques (mentioned in the previous section) can be employed to optimize the parameter values and to study the effects of the time delay in switching encountered physically. A sample of the experimental results is shown in Fig. 2.* The response (output, y, and error, e) of the nonlinear system is compared with the response of the linear system ($\beta_m = \gamma_n = 0$). In both cases the input was a triangle wave whose frequency was varied from $0.1\omega_n$† to $0.5\omega_n$ in $0.1\omega_n$ intervals.‡

The following data are pertinent to these results:

Nonlinear system

$$D = 0.6$$

$$\beta_3 = -\beta_0 = 2 \qquad \gamma_3 = -\gamma_0 = 2$$

$$\beta_2 = -\beta_1 = 0.5 \qquad \gamma_2 = -\gamma_1 = 0.5$$

Linear system

$$D = 0.6$$

Commenting on these results it is noted:

1. The time averages of the magnitude of instantaneous error,

$$\widetilde{|e|} = \frac{1}{T}\int_0^T |e|\, dt \,.$$

compare as follows

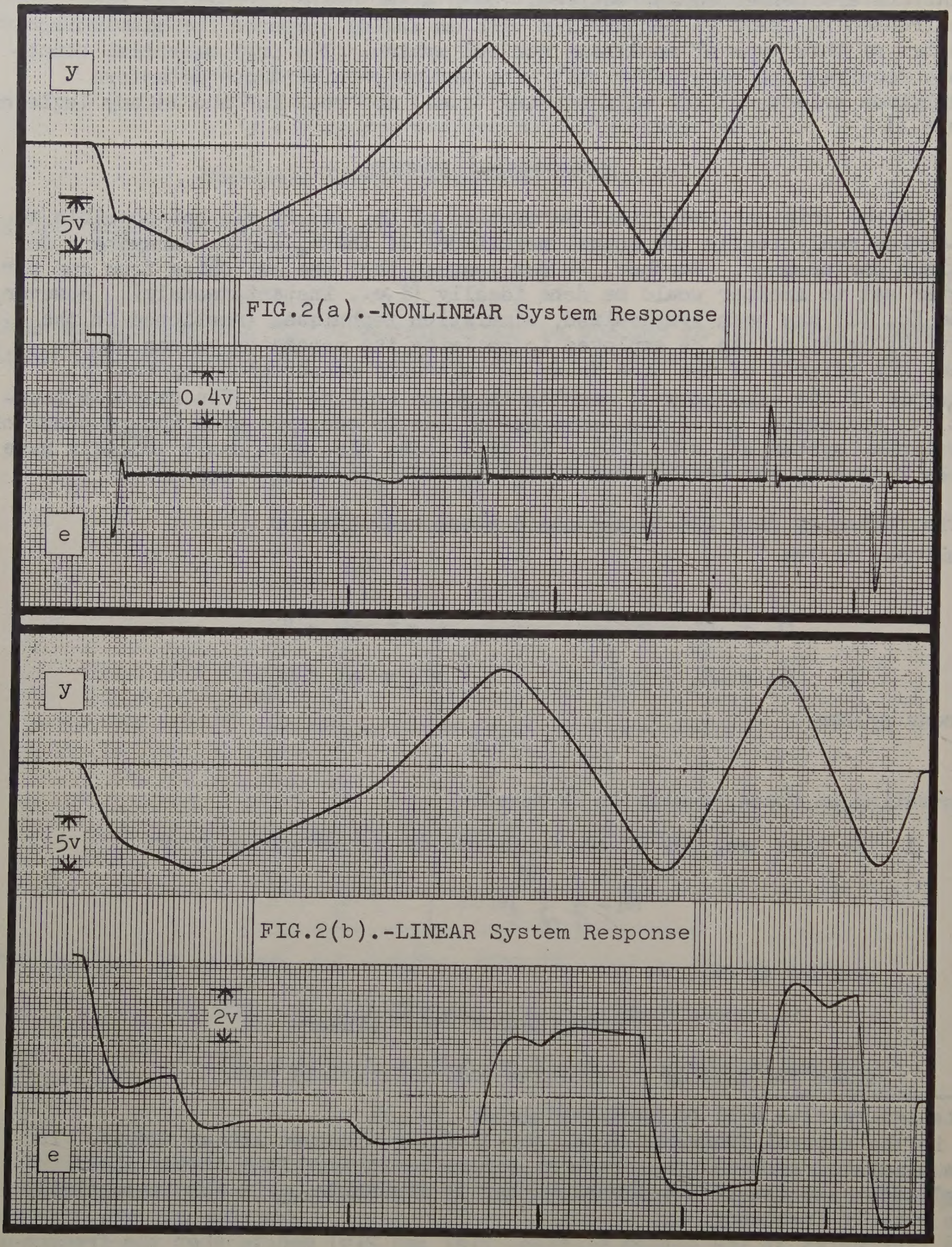
$$\widetilde{|e|}_{\text{nonlinear}} \approx 0.1\, \widetilde{|e|}_{\text{linear}}$$

where the average is over the entire length of run.

*Detailed results will be published later.

†ω_n is the natural frequency (rad./sec.) of the undamped linear member. In the simulated system ω_n has been chosen to be unity. In a steady state amplitude response of the linear member to sinusoids ω_n is also essentially the cut-off frequency, ω_c.

‡The tick marks at the bottom of Figs. 2(a) and 2(b) denote when the frequency was changed. Further, it should be noticed that Figs. 2(a) and 2(b) are two separate sets of results in that they were not obtained simultaneously. Therefore, there is not exact synchronism in the two time scales.



y
5v
FIG.2(a).-NONLINEAR System Response
0.4v
e
y
5v
FIG.2(b).-LINEAR System Response
2v
e

Fig. 2

2. In order to obtain the same improvement in $|\tilde{e}|$ by utilizing linear means alone, the natural frequency of the system should be increased by the same factor as $|\tilde{e}|$ is reduced, i.e., 10 times.

In second order system this requires an increase of loop gain of approximately $10^2 = 100$.

PRELIMINARY EXTENSIONS TO HIGHER ORDER SYSTEMS

From a practical viewpoint limitations in applicability of this type control do not stem from the feedback switching circuit, but rather from assuming that the linear member is second order. In many cases a more realistic approach is to consider the linear member to be of higher order but still predominantly of second order. A third order system of this nature is currently under investigation and promising results are being obtained.

NONLINEAR COMPENSATION OF AN AIRCRAFT INSTRUMENT SERVOMECHANISM BY ANALOG SIMULATION*

D. Lebell
The Ramo-Wooldridge Corporation
Los Angeles, California

Summary -- Combination of certain analog computer techniques with the direct synthesis philosophy[2] applied to a particularly appropriate servo configuration[3] results in a method for direct specification of compensator design.[4] The method centers about a calibration technique which is simple and easily applicable to nonlinear control systems of considerable generality. Mathematical representation of the physical system is maintained at the realistic level typical of analog simulation.

A detailed description of the method is presented by means of its application to a practical example -- nonlinear compensation of an instrument servomechanism. Equations governing the behavior of the unalterable components were written and substantiated experimentally. These equations uniquely determine the form of the ideal compensation for the designated system performance and system configuration. The actual unalterable elements are then employed as an analog simulator to "compute" the specified compensator characteristics. A compensator possessing these characteristic is assembled and over-all system performance obtained. An electronic analog computer is employed[5] as a computer simulator for this work.

The majority of servo compensators designed in the past has been of the linear type. It is well known that this partiality is not due to superiority of the linear over the nonlinear kind but rather that linear designs have been much simpler to perform and nonlinear components hard to come by. The advent of analog computers, the extensive progress made in components development, plus the increasing need for "getting the most" out of available power transducers combine to encourage removal of the highly restrictive linearity condition.

Nonlinear design techniques available to the engineer have frequently suffered from mathematical complexity or computational drudgery on the one hand, or unrealistic simplification on the other. Recently, however, considerable progress has been made in the region between these extremes. This paper attempts to further such progress and its dissemination.

DESCRIPTION OF THE COMPENSATION PROBLEM

An elementary control system representation is shown in Fig. 1. Typically, the designer starts with an element G whose characteristics are fixed by a combination of design specifications, commercial availability and local tradition. His job is to design compensation (or equalization) in order that over-all system performance [response of C(t) to R(t)] be satisfactory. For the instrument servo and for many other applications it is desired that the

*This paper was presented at WESCON, the West Coast Conference sponsored by the IRE - West Coast Electronics Manufacturer's Association, in August, 1955.

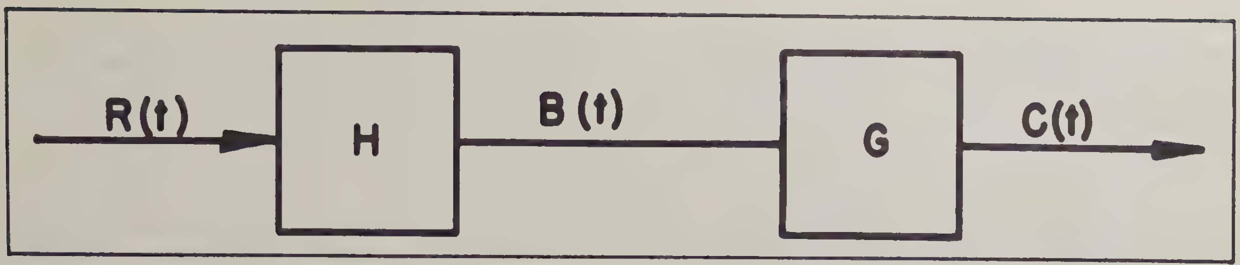


Fig. 1 - Elementary control system.

output and input be identical over a range of inputs. For the output and input to be equal in this configuration, it is necessary for the H component to possess the inverse function characteristics of the G component. For example if $C = G(B) = B^2$, it is necessary that $B = H(R) = \sqrt{R}$, or if $C = \int B dt$, it is necessary that $B = dR/dt$. The extent to which this condition $H = G^{-1}$ can be achieved is the extent to which we can approach the ideal system possessing no discrepancy between R(t) and C(t).

DIRECT SYNTHESIS

The design or synthesis features are exemplified in Fig. 2. Note that R(t) = C(t); G, and consequently B(t) are known. Therefore, the input-output characteristic of H is uniquely determined at least for any particular R(t), and this must be so for any system of any complexity provided only one unknown H exists at a time.

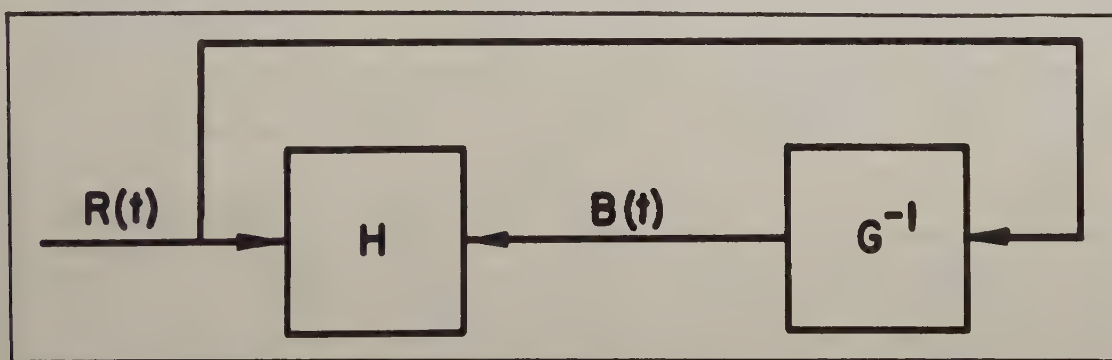


Fig. 2 - Elementary control system with R(t) = C(t) condition imposed.

Our purpose now is to convert the previous statement into a general procedure for obtaining the actual input-output or transference characteristic of H. Figure 2 can be mechanized in a computer set-up as indicated in Fig. 3 where the H position is taken by a recording device (cathode ray oscilloscope) whose vertical and horizontal excursions are controlled as indicated. The pattern traced on the scope is the solution which the computer set-up yields, namely the desired input-output characteristic of H. G^{-1} may be assembled from computer components in accordance with the governing equations. If real-time simulation techniques are employed, G^{-1} may be the actual component.

TECHNIQUES FOR REVERSING SIGNAL DIRECTION THROUGH G

Some comments are certainly in order with respect to the practical problems of reversing the direction of signal flow through G^{-1}. If an analog computer is used to represent G it is not difficult to invert the transfer-

ence. One simply rewrites the differential equation so that the driving function and the dependent variable exchange roles. However, if the actual physical component G is used (which has the advantage of more realistic system representation), its mechanization may present a difficulty.

One approach is to build an auxiliary high-gain servo (it is convenient to use analog computer components here) which compares the actual output of G with the desired output, causing the amplified discrepancy to actuate B and incidentally the cro. This approach is quite satisfactory within the stability and accuracy limitations of this closed loop around G.

A second approach has the advantage of increased simplicity (Fig. 4). A signal is applied at what is normally the input to G [i.e., B(t)]. This

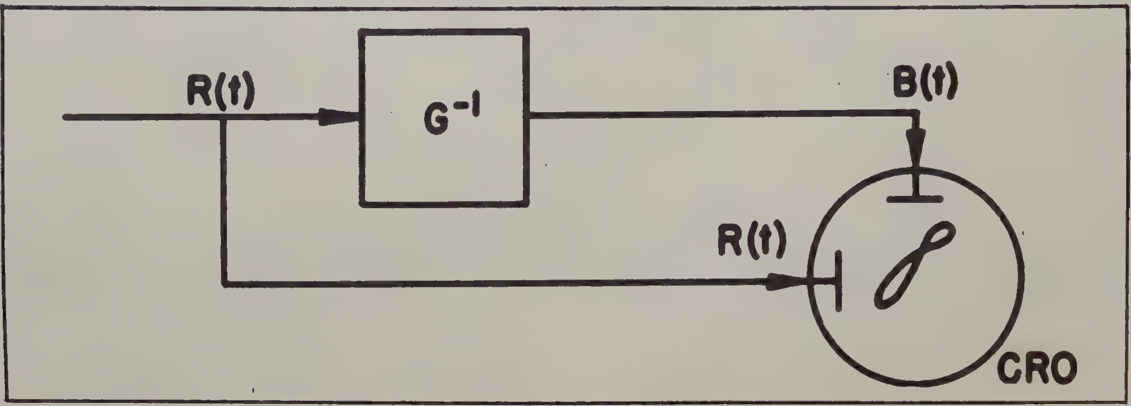


Fig. 3 - Mechanization of Fig. 2 configuration for direct synthesis of compensator (H).

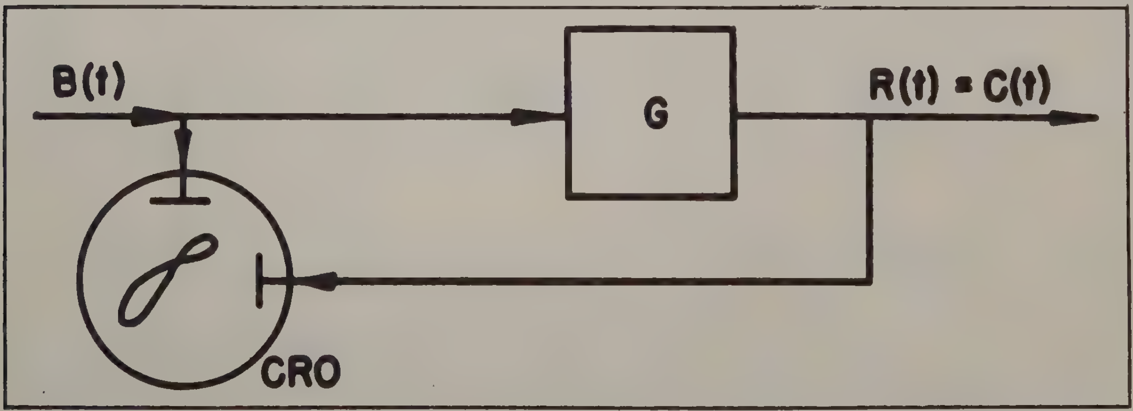


Fig. 4 - Revision of Fig. 3 arrangement to permit normal signal-flow direction through (G).

signal is chosen so that the output of G conforms to the desired range of R. For the moment, G is serving as an analog of itself and also as a signal generator for R(t) which is fed to the cro.

INTERPRETATION OF CRO PATTERN

What are the typical cro patterns one might expect to see? Suppose H is a straight line of slope m. This indicates that H physically should be an amplifier of gain m (modified by scale factors as necessary).

Suppose H is a curved line which stays substantially constant over the signal range. This indicates that a nonlinear transference having this contour as its specific characteristic is required for H. A compensator with this characteristic may be built from nonlinear components such as thyrite, germanium, etc. H may display infinite slope in some regions. This situation frequently occurs as a result of saturation or static friction in G. It constitutes G's legitimate claim on H if the $R(t) = C(t)$ condition is to obtain. If the region is significant it may be approximated by a local high-gain region in the H compensator.

Alternatively, the cro pattern may appear elliptical (on application of a sinusoidal excitation). This is the familiar lissajou figure indicating the need for a linear dynamic characteristic for H.

Finally, a combination of the foregoing may appear, in which case a nonlinear dynamic characteristic is required. For proper interpretation of the cro pattern, however, it must be possible to separate the nonlinear from the dynamic characteristic. One must decompose the H component into subcomponents each of which is linear-dynamic or nonlinear and nondynamic. The correct order of decomposition must be observed if the inverse nature of G and H is to exist. The order in which the H subcomponents are arranged must be inverse to that of G. Strictly speaking, the method is only applicable if no more than one subcomponent is unknown. By judicious choice of signals, however, it may be feasible to apply the method successively to each part of H. For example, a high-frequency low-amplitude R(t) will permit investigation of the linear-dynamic block; a low-frequency high-amplitude signal will isolate the nondynamic nonlinearity. General knowledge of the nature of G will aid inestimably in adapting the method to new situations. Where decomposition of H yields more than two unknown subcomponents the method does not appear to be generally practical.

THE UNALTERABLE ELEMENT G

Conforming to recommendations of the preceding section, preliminary study of unalterable element G is helpful in establishing the form of H to be anticipated. For the particular magnetic amplifier-motor-gear train combination employed the torque equations are:

$$T = K_o(1 - q\dot{C})\,B = J\ddot{C} \qquad (1)^*$$

where

$q = -1$ for $T < 0$
$ = 1$ for $T > 0$
T = developed torque
J = effective moment of inertia of motor and load
$\ddot{C}$ = per-unit motor acceleration
B = amplitude of two-phase 400 cycle voltage
K_o = constant of proportionality
 = 1.0 for suitable choice of units.

*valid for $\frac{dq}{dt} = 0$

Note that developed torque is assumed proportional to the product of slip and control voltage and that representation of slip for bidirectional motor operation requires introduction of the sign operator q. Also, the assumption which more or less defines an instrument servo is invoked, namely, that friction torque is negligible in comparison with inertia torque. Finally, the magnetic amplifier time delay is negligible.

CONFIGURATION FOR H

Equation (1) leads to

$$J\ddot{C} = B(1 - q\dot{C}) = Bq\left(\frac{1}{q} - \dot{C}\right) \qquad (2)^{*}$$

or

$$B = \frac{J\ddot{C}}{q(1/q - \dot{C})} = \frac{-d}{dt}\,\frac{J}{q}\,\ln(q - \dot{C})$$

since

$$q = 1/q$$

In what manner must B be generated from R for the condition (R = C)? Replacing C with R in Eq. (2) results in

$$B = \frac{d}{dt}\,\frac{J}{q}\,\ln(q - \dot{R}) \qquad (3)^{*}$$

which defines the form of H. This relationship is further depicted in the diagram in Fig. 5.

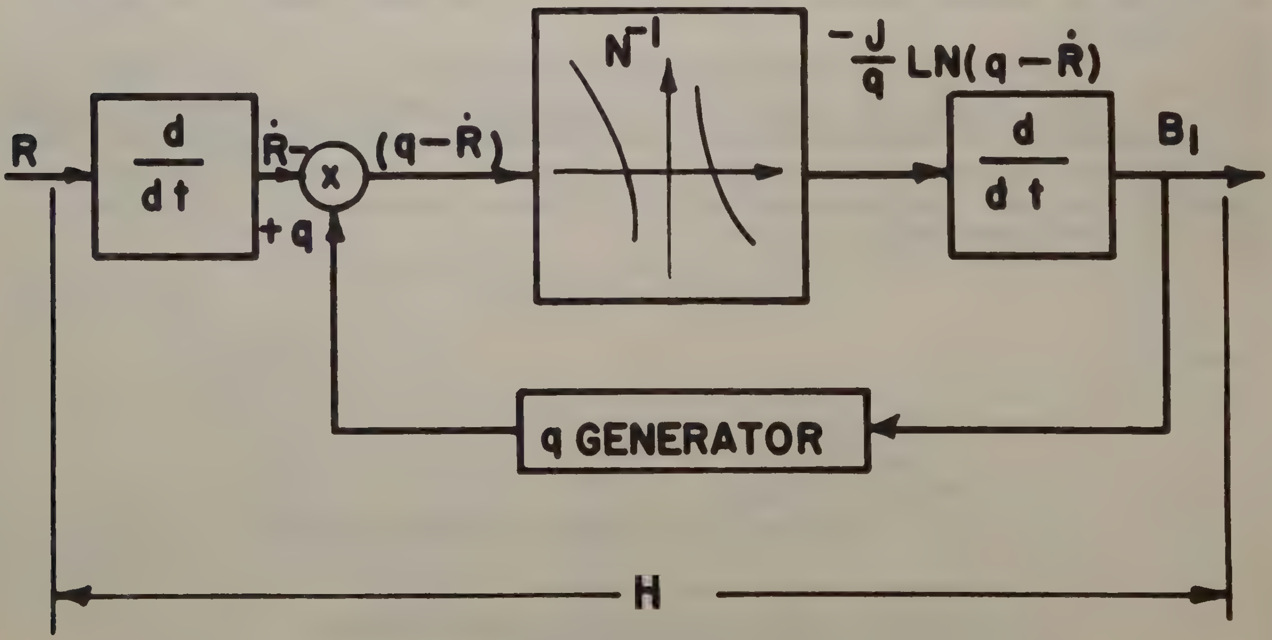


Fig. 5 - Representation of compensator (H) from Eq. (3).

---

*valid for $\frac{dq}{dt} = 0$

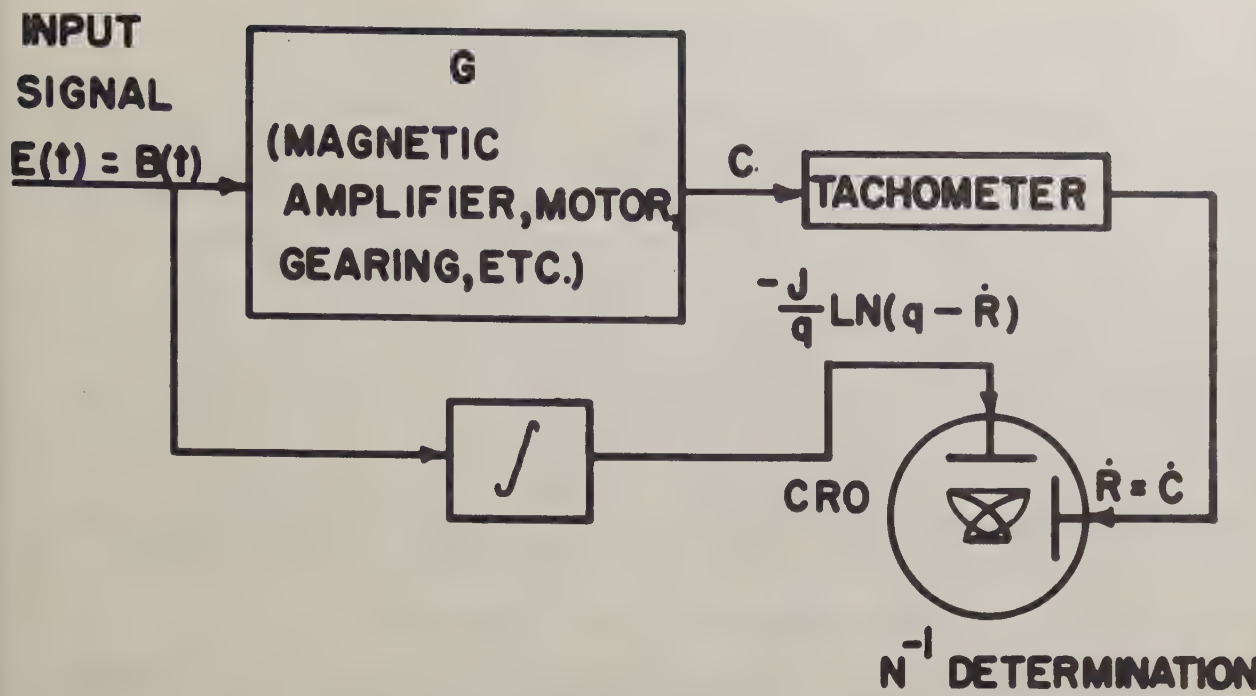


Fig. 6 - Experimental set-up for cro display of nonlinearity.

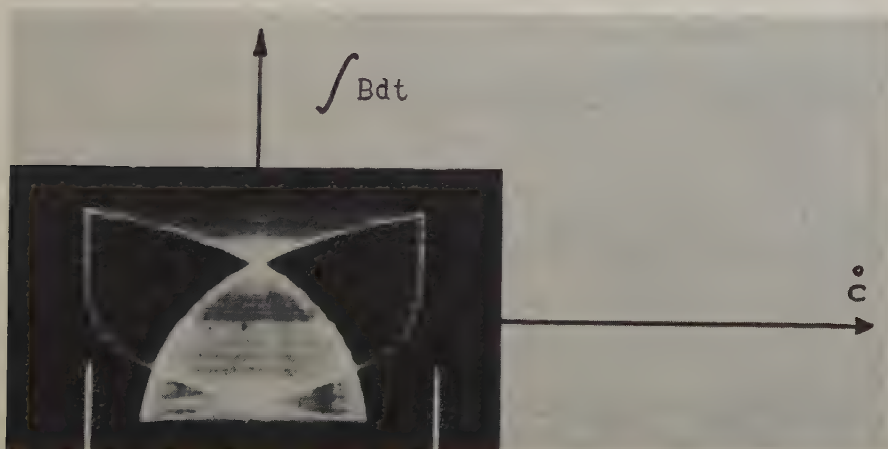


Fig. 7 - Obtained from set-up of Fig. 6 is cro pattern.

EXPERIMENTAL SET-UP

Figure 6 illustrates the arrangement for producing the nonlinear characteristic of the compensator. The motor tachometer serves conveniently as one of the H differentiators. The other is temporarily an integrator since the direction of signal flow is reversed. It is constructed of analog computer elements.

Finally, note the deletion of the q generator. This last simplification produces a piecewise-constant error (known) in the cro pattern which is easily corrected. Theory predicts that the resultant cro pattern should appear piecewise logarithmic which it does (Fig. 7). This pattern was found to be essentially invariant over the operating range of input signals.

INTERPRETATION OF NONLINEAR CRO PATTERN

The multivalued nature of the curve while aesthetically pleasing requires some interpretation. The tachometer output is a 400 cycle voltage whose peak magnitude (envelope) is proportional to speed. For speed reversal, the tachometer voltage reverses phase with respect to line voltage, a phenomenon which is lost in a high persistence cro. From the expected system behavior, however, one can easily select the appropriate half of the envelope. The function is now reduced to a mere double-valued curve. This last ambiguity was to be expected since, as mentioned earlier, the abscissa of the cro pattern is $\dot{R}$ rather than the $q - \dot{R}$ which would have produced a single-valued function. Replotting the curve as a function of $q - \dot{R}$ leads to the N^{-1} characteristic eventually constructed from computer components. Note that we may select an arbitrary horizontal datum for the two branches of the curve since the output of N^{-1} is to be differentiated. This was done to permit the smoothest approximation to the function. In so doing, two biased diodes sufficed to approximate the curve in the final set-up made on the computer.

SYSTEM CONFIGURATION AND PERFORMANCE

The simple configuration in Fig. 1 suffers from the absence of feedback and feed-forward. The configuration in Fig. 8 was chosen to desensitize system performance to component variations. In so doing, it was possible to employ low tolerance components (10 - 25%) and yet obtain the performance indicated in Fig. 9. The data compare performance of the nonlinear, dynamic compensator (Fig. 5) with that of a conventional, lead-lag compensator. The comparison is based on an average error (divided by input) taken over a range of input amplitudes at each frequency for which a data point is indicated.

Several items may be noted with respect to Fig. 9. The sharply rising error (as frequency increases) is the typical "cut-off" phenomenon. An improvement in high frequency performance is attained through a combination of two factors:

1. The nonlinear characteristic in H compensates in part for the nonlinear saturation inherent in G.

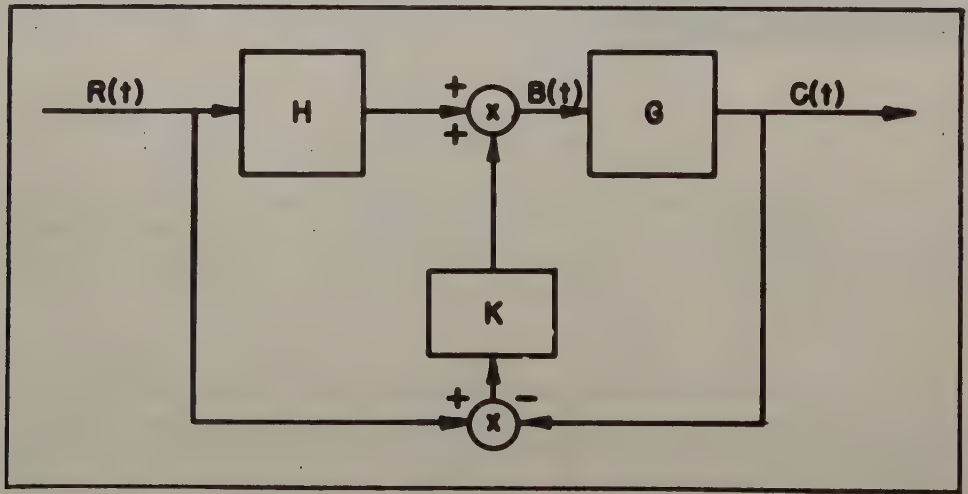


Fig. 8 - Final system configuration employed.

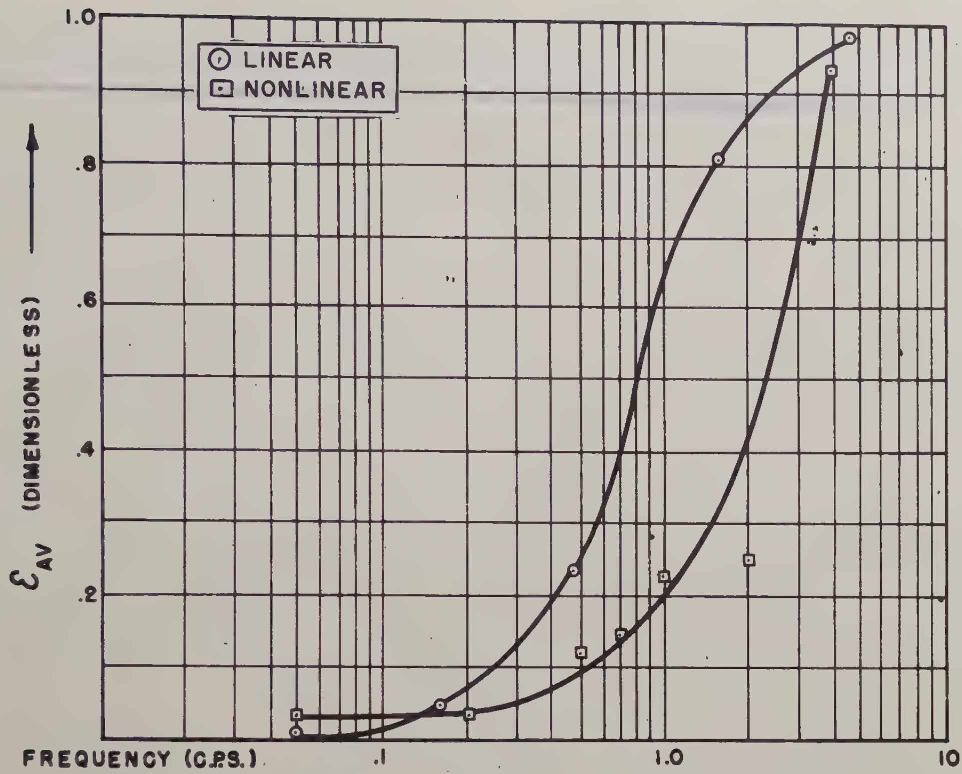


Fig. 9 - Comparison of final system performance with that of conventional lead-lag compensation (direct feedback).

2. The configuration of Fig. 8 permits a greater "lead" or anticipation in H to compensate the "lag" in G. This is due to the stability limitations inherent in the conventional direct feedback configuration.

On the other hand, since the lead-lag compensator permitted closure of the feedback loop at a higher loop gain than in the other case, it displays slightly superior performance in the low frequency region. (Note that H does not pass low frequency signals.) It is apparent that a system employing both types of compensation will possess the attributes of both.

It will be noted from the preceding paragraph that as is frequently the case, improved performance comes at the price of increased complexity. The usual principles of good design demand that a proper balance be struck between complexity and performance for each situation. It should be emphasized that the techniques illustrated are largely independent of the over-all system complexity. They have been tested successfully on other systems of practical interest.

CONCLUSION

Restrictions on the method described are:

1. The unknown compensator (H) must be representable (at worst) by two unknown subcomponents -- nonlinear-nondynamic and linear-dynamic.

2. The cro patterns which result must be capable of synthesis from physically realizable components. Otherwise, H is incorrectly located, or the performance criterion cannot be satisfied.

3. A particular cro pattern is valid for the particular R(t) excitation only. It must remain approximately invariant over the range of anticipated signals if it is to be synthesized from nonlinear-nondynamic elements.

If these conditions can be met, a straightforward calibration method of considerable generality is available wherein the unalterable elements of a system prescribe directly the compensator required to meet the performance specification.

ACKNOWLEDGEMENT

All experimental work described herein was conducted with the aid of analog computer and other equipment facilities of the United States Naval Ordnace Test Station (Inyokern). The helpful cooperation and assistance of I. H. Swift and members of his division of the Aviation Ordance Department is acknowledged.

REFERENCES

1. McDonald, D.: "Multiple Mode Operation of Servomechanisms." Cook Research Laboratories Research Report.

2. Moore, J. R.: "Direct synthesis." Trans. IRE, PGI-1, pp. 6-23 (May 1952).

3. Moore, J. R.: "Combination open-cycle closed-cycle systems." Proc. IRE, 39:1421-1432 (Nov. 1951).

4. Lebell, D.: "Nonlinear Compensation of Control Systems." Doctorate thesis, University of California (Jan. 1955).

5. "Operation and Maintenances of L3 Geda." GER-5164 (June 1, 1953).

6. Hooper, E. T.: "Magnetic Control Amplifiers XM-16A and XM-17A, for Use with Servo Motors Mark 7, 8, 14 and 16." Naval Ordnance Report 2833, U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Md. (May 8, 1953).

7. Kabrisky, M. J.: "The Use of Magnetic Amplifiers to Drive Two-Phase Servomotors." Res. Rep. R-276-52, PIB-215, ONR Contrast Nonr. 292(00) Proj. No. NR-075-215, Polytechnique Institute of Brooklyn, Microwave Research Institute (July 1952).

A STEADY STATE APPROACH TO THE THEORY OF SATURABLE SERVO SYSTEMS

J. C. Lozier
Bell Telephone Laboratories
Whippany, New Jersey

Summary -- It is shown that simple saturable servo systems can have two modes of response to a given input signal -- one a linear mode as predicted by linear feedback theory, and the other a saturated mode predictable from a large signal analysis presented here. This dual mode of response theory provides a reasonable explanation for premature saturation, hysteresis in the input-output characteristics, jump resonance and similar anomalous effects that so often degrade the steady state performance of saturable servo systems. Furthermore, systems which exhibit a large degree of anomalous steady state behavior can be expected to exhibit a correspondingly poor transient response under large signal conditions such as obtained when the system is coming out of saturation.

Demonstration of the existence of this dual mode of response and prediction of the range of signal amplitude and frequencies for which it exists are both built around the "saturated" transfer characteristics of the control loop. This analysis is quantitatively useful only in certain simple cases where the necessary saturated transfer characteristics can readily be found. Its chief value lies in the insight it provides to the cause of anomalous behavior. In this respect it is particularly useful to the designer because it relates this behavior to the loop gain and phase characteristics of the saturated transfer characteristics.

The analysis is applied here to two simple examples, and the results are verified by an analog computer study on simulations of these systems.

The type of saturation discussed here is the familiar power saturation that occurs in the forward path of a linear feedback control system and marks the end of its linear range of operation. In a system where size, weight, efficiency, etc., are important considerations it is often necessary to compromise on the power handling capacity to the point where operation under power saturated conditions is to be expected a good part of the time. Such systems will be referred to here as "saturable" servo systems.

The performance of these systems obviously depends largely on the performance in the neighborhood of saturation. Unfortunately, servo systems often exhibit a large degree of anomalous behavior in this region. This anomalous behavior can affect the steady state performance by causing premature saturation, hysteresis, "jump resonance"[1,2] and similar double valued system response characteristics. Enhancement of the amplitude and lengthening of the period of the overshoot,[3] reduced stability margins in the presence of noise and even low frequency oscillations are transient manifestations of such anomalous behavior.

All the anomalous effects noted above are large signal phenomena. The problem of how a saturable system will perform in the presence of large sig-

nals is a nonlinear one. A linear feedback analysis of the system is helpful to the extent that a satisfactory small signal performance is a prerequisite to satisfactory large signal performance. However, linear concepts are often misleading for handling large signal performance.

Fortunately a form of steady state analysis of these quasi-linear systems can generally be made by fixing the amplitude of the signal at each frequency. This is the basis of Kochenburger's analysis of contactor servos,[4] and it is the basis of this analysis. This approach will be used to derive what is termed here the "saturated" control loop transfer characteristics for these saturable systems. With this saturated transfer characteristic, the possibility of either a linear or a saturated response for the same input signal can be demonstrated. The conditions for the existence of this dual response behavior will be shown to depend on the gain and phase of the saturated transfer characteristics.

The technique is illustrated with two simple examples and the computed results are verified by the use of an analog computer.

DESCRIPTION

In the course of the analysis, two feedback control systems will be studied and compared. They will be referred to here as the "reference" system and the "modified" system. The following assumptions and restrictions concern these systems.

Assumptions and Restrictions

1. Both will be simple single loop systems of the type shown in Fig. 1.

2. In Fig. 1 the box labeled "servo amplifier and equalizer" will comprise the control elements. It will be assumed that the transfer function of this box, G_{23}, will be independent of the amplitude of the input signal.

3. The only difference between the reference system and the modified system is that G_{23} is just a constant factor K in the case of the reference system, whereas G_{23} will have a specified amplitude and phase vs. frequency characteristic, as well as a gain factor, in the case of the modified system.

4. The only nonlinearity in these systems will be in the box labeled "limiter." This limiter will have an output amplitude proportional to its input amplitude up to the limiting point, as shown in Fig. 2. The limiter is assumed to have no amplitude or phase vs. frequency characteristics, so that its transfer function G_{34} is just a constant gain factor below limiting and a variable gain factor above.

5. The box labeled "servo motor and load" will be assumed to have a linear input-output characteristic for all input amplitudes passed by the limter. In other words, the saturation takes place in the limiter just ahead of the motor.

6. The steady state transfer characteristics of the motor and load, G_{45}, is assumed to be that as shown in Fig. 3.

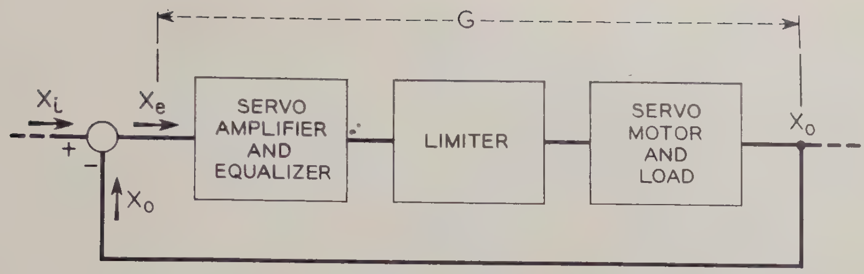


Fig. 1

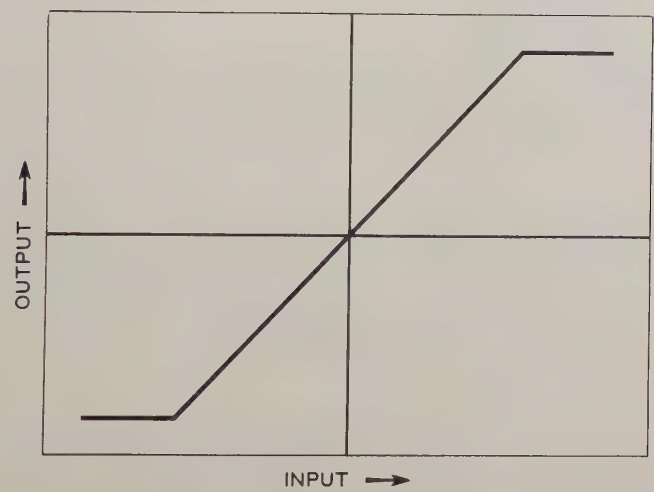


Fig. 2

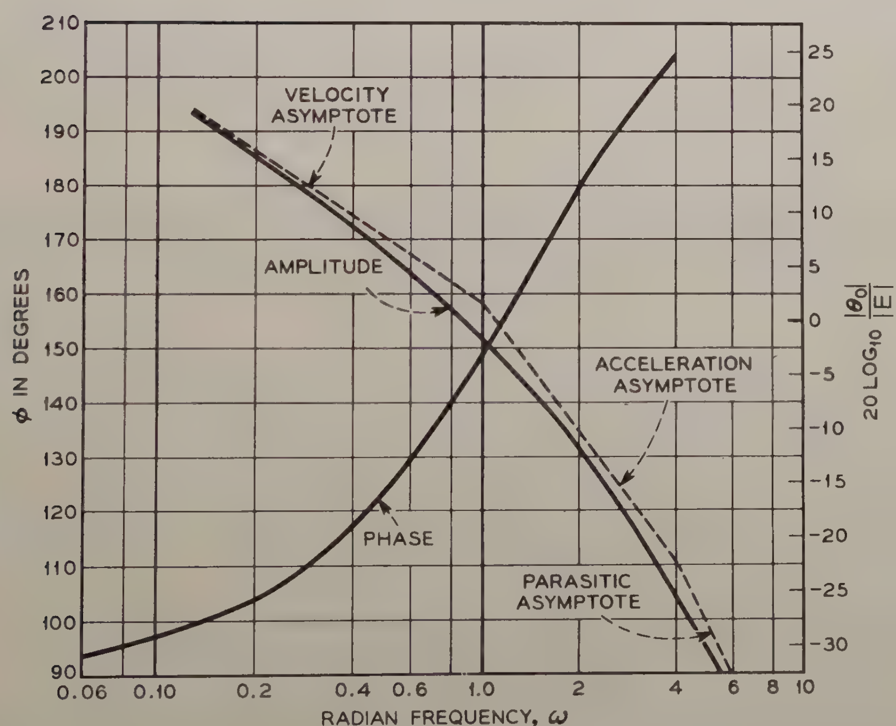


Fig. 3

Note: The shape of this transfer function is fairly typical of such motors. The constant velocity and constant acceleration asymptotes which bound the output displacement vs. frequency characteristic over much of its useful frequency band follow directly from the assumption that the velocity and the torque the motor can develop are proportional to the amplitude of the input voltage. A parasitic asymptote such as the one assumed here is also typical of such motors. It was added for completeness and has little bearing on the main arguments given here. The associated phase vs. frequency characteristic assumed here is simply the minimal* phase that must accompany the assumed amplitude vs. frequency characteristic.

7. It will be further assumed that the shape of the displacement vs. frequency, and of the phase vs. frequency characteristics of the motor are the same whether the motor is driven by a sine wave signal or by a square wave of the same frequency.

Note: This assumption ignores the harmonics of the fundamental component that may be present in either the input or output signals. Such assumptions are customarily justified in computations of this type on the basis that feedback control systems are low pass devices. One exception to this will be made below.

8. It is assumed that neither the relative shape of the displacement vs. frequency characteristic nor the actual phase vs. frequency characteristic of the limiter plus the motor is changed by saturation. In other words, saturation in the limiter and motor combination is assumed to just change the flat gain factor in the limiter transfer function.

These assumptions have been chosen to make it easy to derive the steady state saturable transfer characteristics for the limiter plus motor and load.

Derivation of the Saturated Transfer Characteristics

The maximum output displacement vs. frequency characteristics of the limiter plus motor and load characteristics may readily be derived from the linear transfer characteristics shown in Fig. 3, using the above assumptions. In the linear case, for an input voltage to the limiter of the form

$$e(t) = E \cos wt$$

the output displacement of the motor and load is

$$\theta_o(t) = |\theta_o| \cos (wt + \phi)$$

and Fig. 3 shows curves of $20 \log_{10} \frac{|\theta_o|}{|E|}$ and ϕ as functions of frequency.

In saturation the output contains a fundamental component

$$|(\theta_o)_m| \cos (wt + \phi)$$

*Bode, H. W.: "Network Analysis and Feedback Amplifier Design." D. Van Nostrand Company.

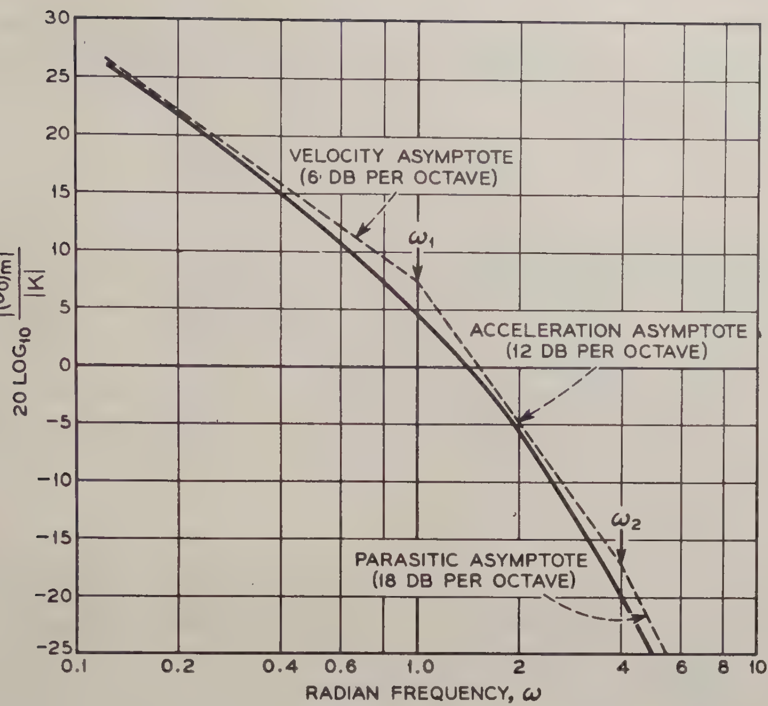


Fig. 4

with $|(\theta_o)_m|$ constant at any given frequency for any input voltage E exceeding the saturation value E_s. Figure 4 shows a plot of $20 \log_{10} \frac{|(\theta_o)_m|}{K}$ as a function of frequency, where K is a constant. This maximum output displacement vs. frequency characteristic has the same relative shape as the linear characteristic due to assumptions 4, 5, 6 and 7 above. The saturated phase characteristic is identical to the linear phase characteristics again because of assumptions 4, 5, 6 and 7.

Based on these simplifying assumptions and restrictions, the desired steady state saturated transfer function for limiter and the motor plus load can be expressed as

$$(\theta_o)_m(w) = G_{35}(w)E_s. \tag{1.1}$$

Here $(\theta_o)_m$ is the maximum output displacement vs. frequency, and E_s is the input to the limiter that will just produce saturation. One major difference between the saturated transfer function given in Eq. (1.1) and its linear counterpart is that this saturated transfer function only holds for one single input frequency* at a time. This follows from the fact that the superposition theorem used in linear analysis to handle multiple input signals breaks down in this case.

In many saturable transfer elements, both the shape of the amplitude vs. frequency characteristic, and the phase vs. frequency characteristics will change with the amplitude of the saturating signal. In these cases, of course, the saturated transfer characteristics would be different from the

*By assumption 7, repetitive square waves or triangular waves are included.

linear characteristics. In fact, there may actually be a whole family of such saturated transfer characteristics, depending on the precise amplitude of the input signals. Such factors would complicate application of the steady state analysis presented below.

STEADY STATE ANALYSIS

The first part of this section will be devoted to showing the basic techniques used here. The effects of output saturation in producing hysteresis in the steady state input-output characteristics will be demonstrated, and the equations governing the relationship between hysteresis and the saturated loop transfer characteristics will be derived.

In the second part of this section, the technique will be applied to two illustrative examples.

The Basic Technique

The technique used here is based on the relationship that must exist between the three signals associated with the linear summing point in the servo shown in Fig. 1. These signals are the loop input signal $x_i(t)$, the loop feedback signal $x_o(t)$ and the difference or error signal $x_e(t)$. Since x_e is the difference between x_i and x_o, the equation governing this relationship is simply

$$x_e(t) = x_i(t) - x_o(t). \tag{1}$$

Now for a single frequency input, $x_i(t) = A_i \cos(wt)$, $x_e(t)$ and $x_o(t)$ are also sinusoids of the same frequency, ignoring harmonics generated by the saturation in the output. Hence, we can transform Eq. (1), which is good at any instant of time, into the vector equation,

$$X_e(w) = X_i(w) - X_o(w) \tag{2}$$

which holds good at any frequency w.

According to Eq. (2), these three vectors must form a triangle at any frequency w. The triangle that is most useful for illustrative purposes is that shown in Fig. 5. In drawing the triangle in this way, X_e rather than X_i

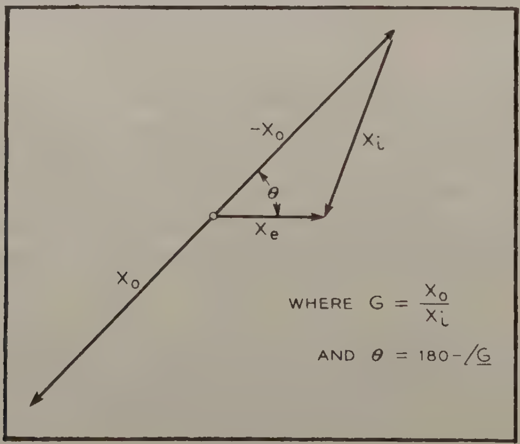


Fig. 5

is the reference vector. This is convenient because X_e and X_o are related by the loop transfer function G, so that

$$X_o(w) = G(w)\ X_e(w). \tag{3}$$

Hence, with X_e as the reference vector, $-X_o$ can be readily determined from G. Once X_e and $-X_o$ are drawn as shown, then X_i must be the vector connecting $-X_o$ and X_e. Incidentally, this method of representing X_e and $-X_o$ is related to the one used in drawing the Nyquist diagram of a system. It should be noted that the angle between X_e and $-X_o$ is $180^o - \underline{/G}$ and therefore represents the phase margin of the control loop transmission at the frequency in question. The symbol θ will be used for this phase margin, so we have

$$\theta = 180^o - \underline{/G}. \tag{4}$$

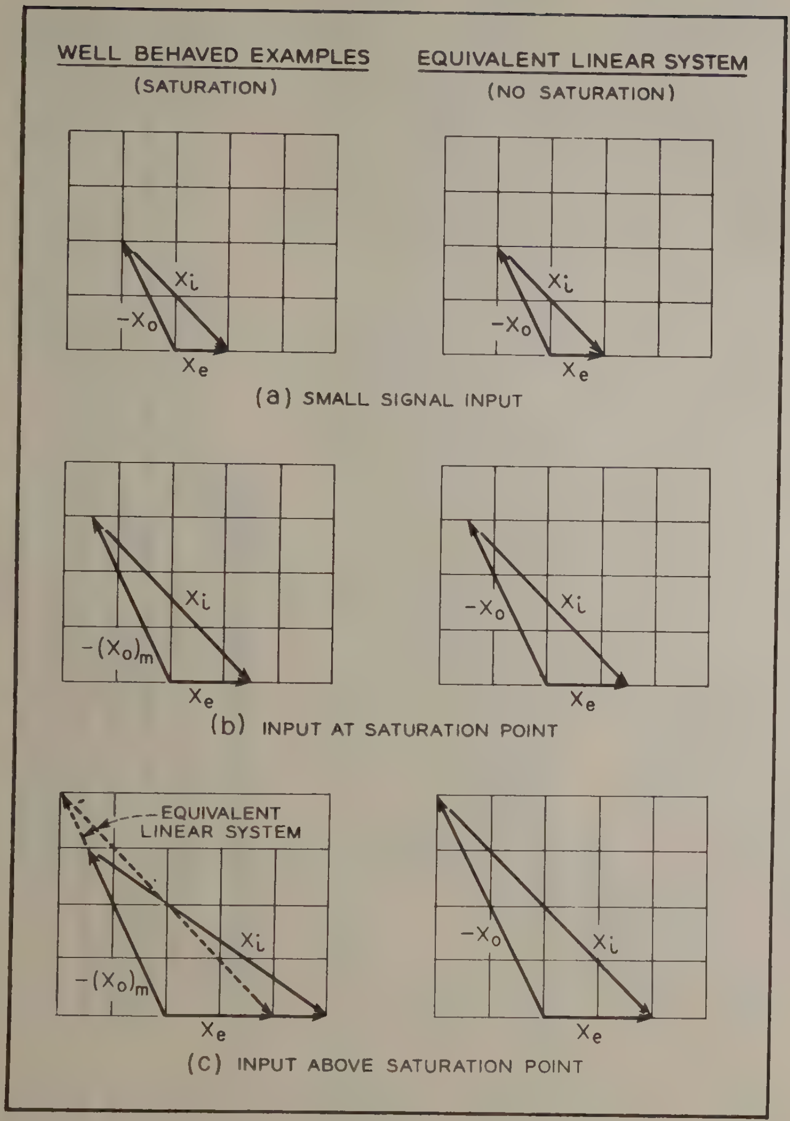
Introduction of Output Saturation

The main interest here centers on what happens as X_o saturates. From the assumptions made earlier, it follows that X_o reaches a maximum value $(X_o)_m$ when X_e reaches the value $(X_e)_s$. X_e is free to increase beyond the saturation value $(X_e)_s$, but $-(X_o)_m$ and $\underline{/G_s}$, the angle between X_e and $-(X_o)_m$, do not change. Hence, as the input X_i is increased beyond $(X_i)_s$, the level that produces saturation, the magnitude of X_e must change in order to complete the triangle of X_e, $-(X_o)_m$ and X_i. Figure 6 shows this phenomenon graphically. In Case (a) the input is below saturation; in Case (b) the input is at saturation level; in Case (c) the input is above saturation. For comparison purposes, Fig. 6 also shows what would happen in the equivalent linear case, where no saturation occurs. As Case (c) shows, saturation prevents $-X_o$ from increasing beyond $-(X_o)_m$, hence X_e, the error, must increase as X_i increases to maintain the vector relationship between $-(X_o)_m$, X_i and X_e. This example in Fig. 6 shows what happens in a well behaved case of saturation, where the relationship between X_i and $-X_o$ is single valued, and all three vectors can be conclusively determined.

The example shown in Fig. 7 is one where the large signal behavior is not so well behaved. We see here the same three input conditions that were used in Fig. 6, and the equivalent linear case is again drawn for reference purposes. In Case (a), where the input is small, the system is effectively linear as before. In Case (c), we again have no trouble determining the vector relationship. However, in Case (b), where the input level is just enough to produce saturation, two possible angular positions are shown for X_i, and two possible magnitudes for X_e.

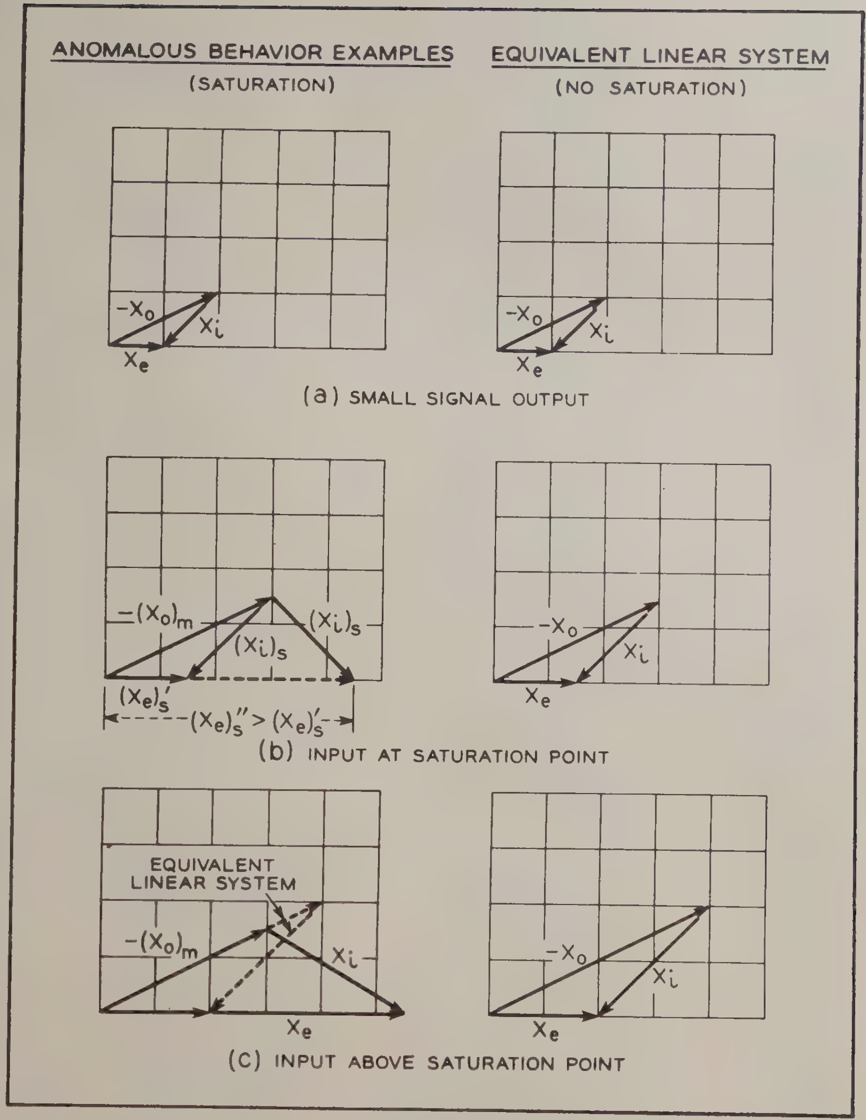
Conditions for Anomalous Large Signal Behavior

Further examination of the example in Fig. 7 will show that $(X_i)_s$ is not the only input amplitude for which the input-output characteristics of the loop has double valued possibilities. Another case is shown in Fig. 8. Here the loop characteristics are again the same as those for Fig. 7, although the scale factor of the drawing has been expanded for clarity. The input amplitude shown here is intermediate between cases (a) and (b) shown in Fig. 7.



WELL BEHAVED EXAMPLES
(SATURATION)
EQUIVALENT LINEAR SYSTEM
(NO SATURATION)
X_i
$-X_o$
X_e
(a) SMALL SIGNAL INPUT
$-(X_o)_m$
(b) INPUT AT SATURATION POINT
EQUIVALENT LINEAR SYSTEM
(C) INPUT ABOVE SATURATION POINT

Fig. 6



ANOMALOUS BEHAVIOR EXAMPLES
(SATURATION)
EQUIVALENT LINEAR SYSTEM
(NO SATURATION)
$-X_o$
X_i
X_e
(a) SMALL SIGNAL OUTPUT
$-(X_o)_m$
$(X_i)_s$
$(X_e)_s'$
$(X_e)_s'' > (X_e)_s'$
(b) INPUT AT SATURATION POINT
EQUIVALENT LINEAR SYSTEM
(C) INPUT ABOVE SATURATION POINT

Fig. 7

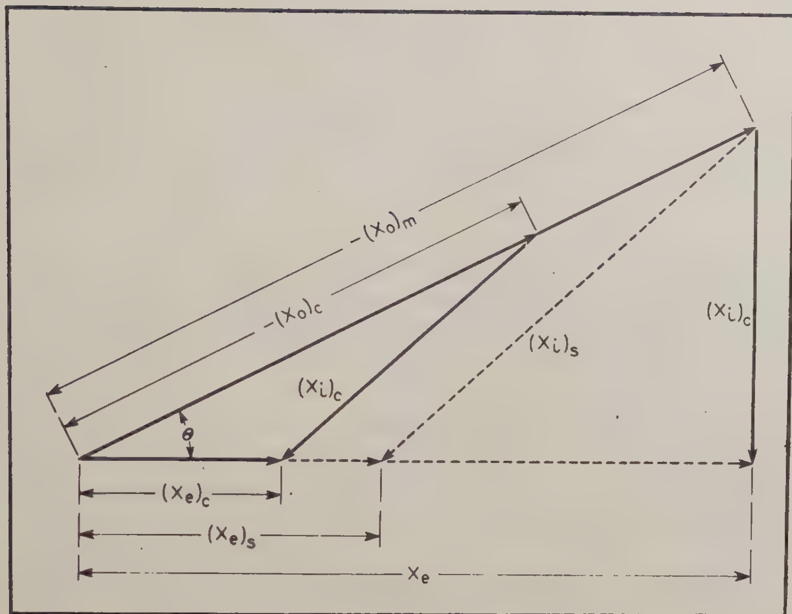


Fig. 8

The precise input amplitude used in Fig. 8 is designated $(X_i)_c$ and will be called the "critical" value. The analysis below will show why this critical value is of particular interest.

As Fig. 8 shows $|(X_i)_c|$ is the magnitude of the perpendicular from $(X_o)_m$ to the X_e axis. Hence, we can write

$$|(X_i)_c| = |(X_o)_m| \sin \theta. \tag{5}$$

The importance of this critical input amplitude stems largely from the fact that it is the smallest amplitude of X_i that could possibly form a closed triangle with X_e, θ and $(X_o)_m$ as shown. Hence, no input amplitude $|X_i| < |(X_i)_c|$ could produce saturation, and the system must be linear for these smaller amplitudes.

However, when $|X_i| = |(X_i)_c|$ the system can have two possible responses; one is the linear response $(X_o)_c$ and the other is the saturated response $(X_o)_m$. The linear response is obviously a stable operating condition since $|(X_i)_c| < |(X_i)_s|$. As for the saturated condition, $(X_i)_c$ can form the necessary triangle with $(X_o)_m$, X_e and θ as shown in Fig. 8. The question of whether this triangle represents a stable operating condition depends on whether or not the magnitude of X_e involved is large enough to sustain the output at the saturation level $(X_o)_m$. Now the magnitude of X_e involved is readily seen to be

$$|X_e| = |(X_o)_m| \cos \theta. \tag{6}$$

The minimum amplitude of X_e that can sustain the output at saturation is $(X_e)_s$. This triangle, therefore, will represent a stable operating condition when

$$|(X_e)_s| < |(X_o)_m| \cos \theta. \tag{7}$$

In the example shown in Fig. 8 this inequality is satisfied, and the second triangle with $|X_o| = |X_o|_m$ is a possible operating condition.

Thus when Eq. (7) is satisfied, the system has two possible equilibrium conditions for the same input amplitude $|X_i| = |(X_i)_c|$. This represents a type of anomalous behavior for which the output response is either the linear value corresponding to the input amplitude or the saturated value $(X_o)_m$.

Similar arguments can be presented to show that such anomalous large signal behavior must exist for all input amplitude between this critical input amplitude $|(X_i)_c|$ and $|(X_i)_s|$, the input amplitude for which saturation is to be expected.

The conditions for anomalous large signal behavior at saturation expressed by the inequality in Eq. (7) deserve further discussion. It was assumed that the system under discussion was linear for all error signals $|(X_e)| \leq |(X_e)_s|$. Hence, we write

$$(X_o)_m = G(X_e)_s. \tag{8}$$

When Eq. (8) is combined with Eq. (7), we see that $|(X_e)_s| < |G||(X_e)_s| \cos\theta$ or, cancelling $(X_e)_s$, that

$$|G| \cos\theta > 1. \tag{9}$$

This equation expresses the conditions of anomalous steady state behavior in terms of the loop transfer characteristics.

<u>Effect of Saturation on the System Transfer Characteristics</u>

The steady state operation of a simple linear servo system can generally be described by three system transfer functions relating the three variables X_i, X_e and X_o. These three transfer functions are:

1. The steady state input-output transfer function, relating X_o to X_i, which can be written for the linear case as

$$\frac{X_o}{X_i} = \frac{G}{1+G}. \tag{10}$$

2. The steady state input to error function, relating X_e to X_i, can be written for the linear case as

$$\frac{X_e}{X_i} = \frac{1}{1+G}. \tag{11}$$

3. The loop transfer function G relating X_o to X_e can be written for the linear case as

$$\frac{X_o}{X_e} = G. \tag{12}$$

The large signal analysis will now be used to determine how these three linear transfer functions are modified when saturation is present. As the previous analysis shows there are two conditions of interest.

The first condition is the one for which the loop transfer characteristics are such that

$$|G| \cos \theta < 1. \tag{13}$$

In this case, no anomalous large signal will occur, and the linear Eqs. (10), (11) and (12) still hold for the three transfer functions. However, their applicability must be restricted to input amplitudes $|X_i| < |(X_i)_s|$ because of saturation.

The second condition is one for which

$$|G| \cos \theta > 1. \tag{9}$$

In this case, the analysis shows that anomalous behavior will occur for input amplitudes above the critical amplitude, $|X_i|_c$, where

$$|X_i|_c = |(X_o)_m| \sin \theta. \tag{5}$$

Hence, there are two subsidiary cases depending on the input amplitude.

Case a. For $|X_i| < |(X_i)_c|$, the linear equations still hold for the three loop transfer functions, as in the first case above, except that the restriction on input amplitudes is determined by $(X_i)_c$ instead of $(X_i)_s$.

Case b. For inputs where such anomalous large signal behavior occurs, namely for (X_i) so that $|(X_i)_c| < |X_i| < |(X_i)_s|$, we have a new set of transfer functions.

(1) The input-output transfer characteristics, relating X_o to X_i, is now double valued with the output having either the value given by the linear Eq. ~~(10)~~ or the saturated value $|(X_o)_m|$. Hence, we have the anomalous situation where

$$\frac{X_o}{X_i} = \frac{G}{1 + G},$$

or (14)

$$\frac{|X_o|}{|X_i|} = \frac{|(X_o)_m|}{|X_i|}.^*$$

(2) The steady state input to error relating X_e to X_c is also double valued in this case. It can have the linear value

$$\frac{X_e}{X_i} = \frac{1}{1 + G}$$

*The angle between X_o and X_i is ambiguous.

or, when the output is saturated,

$$\frac{|X_e|}{|X_i|} = \frac{|(X_i)_c|}{|(X_i)|} \cot\theta \sqrt{1 - \frac{|(X_i)_c|^2}{|(X_i)|^2}}\ ^{*}. \tag{15}$$

(3) The loop transfer function relating X_o to X_e is unaffected by this anomalous behavior. The linear relationship given by Eq. (12) still applies for amplitudes of $|X_e| < |(X_e)_s|$.

EFFECT OF SATURATION ON THE STEADY STATE PERFORMANCE

Using the system transfer characteristics developed in the previous section, the effect of saturation on the steady state performance can now be discussed.

The steady state input-output transfer function will exhibit anomalous large signal behavior unless

$$|G| \cos\theta < 1. \tag{13}$$

This means that the phase margin would have to approach 90° when G, the loop gain, is large, in order to avoid anomalous behavior. Such a restriction is too severe for most practical servo systems to meet at all frequencies, when other requirements such as bandwidth or speed of response are considered. Hence, anomalous large signal behavior is to be expected over some part of the frequency range of the system.

The type of input-output response that is obtained under anomalous conditions is shown in Fig. 9. It shows the typical double valued output that occurs for input amplitudes between $|(X_i)_c|$ and $|(X_i)_s|$. This double valued effect similar to hysteresis and is certainly not desirable. It is apparent that the dependable range for linear operation ends at $X_i = (X_i)_c$.

The effectiveness of the system in utilizing the available power handling capacity under these conditions is measured by the ratio of $|(X_i)_c|$ to $|(X_i)_s|$, or still better by the ratio of $|(X_i)_c|$ to $|(X_o)_m|$. By rearranging Eq. (5) we get

$$\frac{|(X_i)_c|}{|(X_o)_m|} = \sin\theta. \tag{16}$$

---

*The derivation of this expression is straightforward but lengthy. It is based on the fact that

$$\frac{|X_e|}{|(X_i)_c|} = \frac{|(X_o)_m| \cos\theta}{|(X_o)_m| \sin\theta} = \cot\theta.$$

The phase is again ambiguous. A negative sign in front of the radical in Eq. (15) is also possible over a small range of values of X_i, but it seems unlikely from physical considerations.

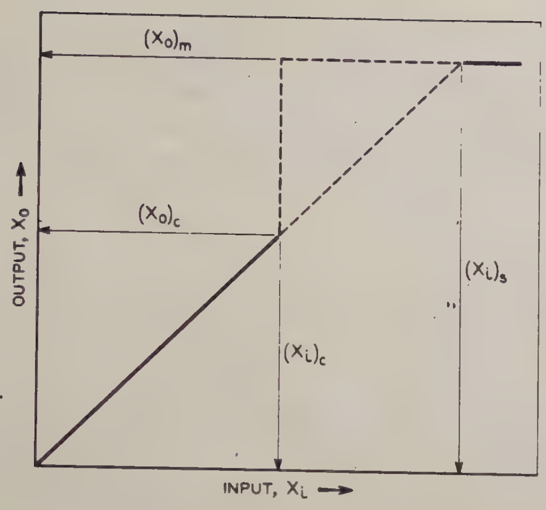


Fig. 9

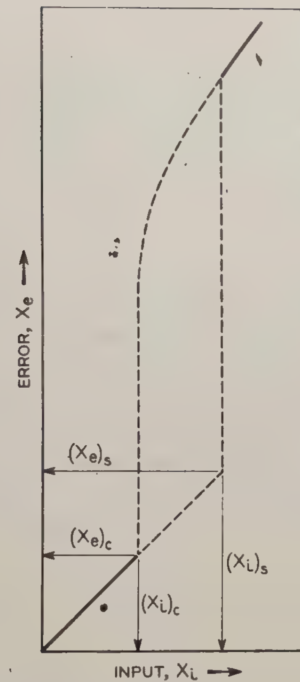


Fig. 10

These results indicate, for example, that at a frequency for which the phase margin is 30^o and the loop gain $G > 1$, the input-output response of the servo is single valued and linear over only half of its potential range. The importance of keeping θ as large as possible over the working frequency range of the system is apparent.

The steady state input to error transfer function of the system also exhibits large signal anomalous behavior under the same conditions as the i put-output transfer functions. Figure 10 gives an example of such behavior. A case of particular interest is that for which $|X_i| = |(X_i)_c|$. The corresponding magnitude of X_e can have either the linear value

$$|X_e| = |(X_e)_c| = \frac{|(X_i)_c|}{|1 \quad G|}, \tag{17}$$

or

$$|X_e| = |(X_i)_c| \cot \theta. \tag{18}$$

It may be shown that the larger the phase margin θ at the frequency in question, the smaller the difference that will exist between the two values for $|X_e|$, as given by Eq. (17) and (18). Similarly, it can be shown that for the same phase margin θ, this difference increases as $|G|$ increases. This can be seen from a comparison of Eqs. (16) and (18). Again the importance of keeping θ large when $|G|$ is large is emphasized.

APPLICATION OF STEADY STATE ANALYSIS

In this section we will analyze and compare the steady state performance of two systems which will be referred to as the reference and the modified systems. These examples have been chosen in order to demonstrate the conclusion that in saturable servo systems the loop phase margin is an important

factor in the performance, even at frequencies where the loop gain $|G| \gg 1$. The reference servo has what may be termed a fairly satisfactory loop phase margin under high loop gain conditions, and the modified servo has a relatively poor one. In other respects the two systems are as much alike as possible. To make the analysis simple, the seven assumptions made previously will apply to both systems. Both will be simple single loop systems as shown in Fig. 1.

The first step will be to present the loop transfer functions for these two systems. In the case of the reference servo, it is assumed that the control element, representing the servo amplifier and equalizer, has a constant nonfrequency sensitive transfer function. Hence the linear loop transfer characteristics are the same as the characteristics shown in Fig. 3 for the limiter plus the motor and load, except for a constant gain factor. The resultant loop transfer characteristics for the reference servo are shown in Fig. 11. The absolute gain around the loop in this case was chosen so that the gain margin at phase crossover ($\theta = 0$) is about 17 db.

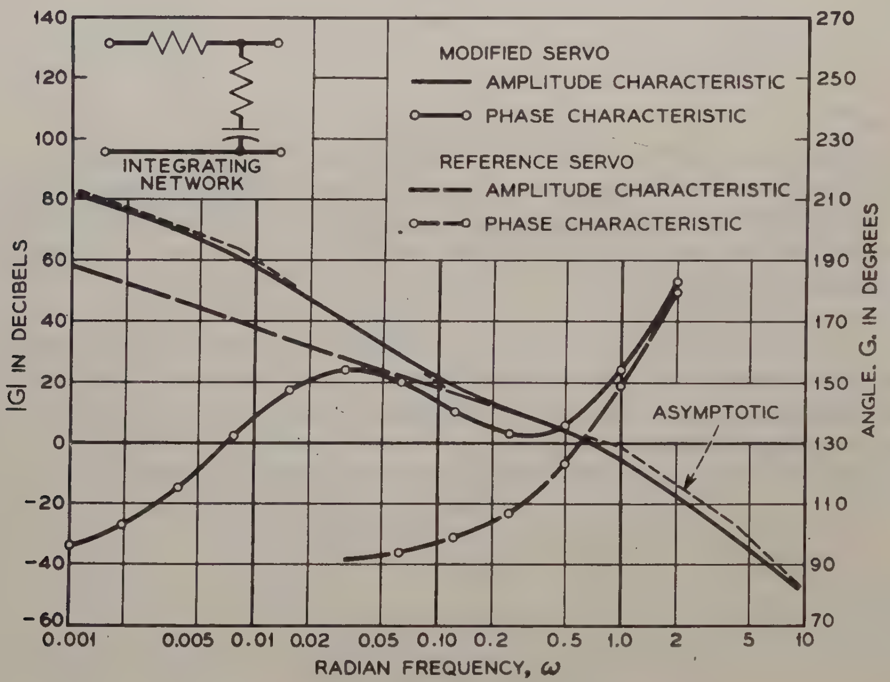


Fig. 11

The linear transfer characteristics for the modified servo were obtained by adding an integrating network of the type shown in Fig. 11 to the reference servo, and then readjusting the flat loop gain until the gain margin at phase crossover is again about 17 db. The resultant loop transfer characteristics are also shown in Fig. 11. The constants of the integrating network were chosen so as to increase the relative loop gain at frequencies below 0.0078 by 24 db with the effect tapering off to zero at frequencies above 0.25 radians per second, as shown.

Since both the reference and the modified servos use the same limiter plus motor and load, the saturated output amplitude vs. frequency characteristic given in Fig. 4 holds for both systems.

Input-Output Performance

Both of these systems were found to satisfy the conditions for anomalous large signal behavior for all frequencies of interest below approximately 0.6 radians per second. Hence, to determine the effect of saturation in the input-output behavior for these systems, we computed $\frac{|(X_i)_c|}{|(X_o)_m|}$ over this frequency range below 0.6 radians per second. Equation (16), which shows that $\frac{|(X_i)_c|}{|(X_o)_m|} = \sin\theta$, was used in this case. The necessary values of $|(X_o)_m|$ used in the computations are given in Fig. 4 for both systems. The phase margin θ in each case was derived from the linear transfer characteristics in Fig. 11. This convenient procedure is allowable on the assumption that θ does not change with saturation in these two systems. The results of the computations are both given in Fig. 12. The curve for the reference system shows that $\frac{|(X_i)_c|}{|(X_o)_m|}$ is very nearly unity except for frequencies in the neighborhood of $w = 0.6$ radians per second. This means that the range of input amplitudes exhibiting anomalous behaviors is negligibly small, and the system will be linear for input amplitudes practically up to $|(X_o)_m|$. On the other hand, the curve for the modified servo shows that this ratio drops as low as -7 db at $w = 0.03$ radians per second, and does not approach the ideal value of unity until w drops below 0.002 radians per second. This means that the anomalous behavior exists for over half of the nominally linear range of input amplitudes for frequencies around $w = 0.03$ radians per second. Hence, the input-output characteristics ceases to be linear when the output is still 7 db below its saturation value at this frequency.

Compared with the reference servo, this is poor behavior. This poor behavior must be due to the reduction in loop phase margin, resulting from the introduction of the integrating network. The only other difference between the two systems is in loop gain at the low frequencies, and gain does not affect this ratio, as Eq. (16) shows.

Figure 12 also shows a plot of the ratio $\frac{|(X_i)_s|}{|(X_o)_m|}$ computer on the assumption that the modified servo behaves linearly for input amplitudes below saturation. This would be the performance of the system if there were no anomalous large signal behavior.

A similar comparison of $\frac{|(X_i)_s|}{|(X_o)_m|}$ and $\frac{|(X_i)_c|}{|(X_o)_m|}$ for the reference servo showed them to be practically identical, which is not surprising in this case where the anomalous effect is very small.

The name "jump resonance" is given to a discontinuity that may occur when the input vs. output characteristic of a servo is measured by a sinusoidal input signal of constant amplitude that sweeps back and forth over the frequency range of interest. The phenomenon is distinguished by a hysteresis effect in the output caused by a sudden jump from the linear to the saturated

output response at a relatively high frequency as the sweep frequency is increasing, and a continuance of the output at the saturation level during the return sweep until, at some lower frequency, the output falls back to the linear value. This phenomenon occurs only when the input amplitude falls within certain limits. Such a hysteresis effect in the input-output characteristic with frequency could occur in the case of the modified servo where $\frac{|(X_i)_c|}{|(X_o)_m|}$ is small, because a signal amplitude that equals the linear saturation value $|(X_i)_s|$ at one frequency would not fall below $|(X_i)_c|$ until a lower frequency was reached on the reverse frequency sweep. Under such conditions, a hysteresis of the type described is quite reasonable.

Input-Output Measurements

The steady state results on these two systems can be checked, so an analog computer was set up to simulate these two systems. Then the output amplitude was measured as a function of the input amplitude. The input frequency used was 0.0316 radians per second. The reference servo showed no measurable anomalous behavior at this frequency. The modified servo gave the results shown in Fig. 13. This measured input-output characteristic is dual

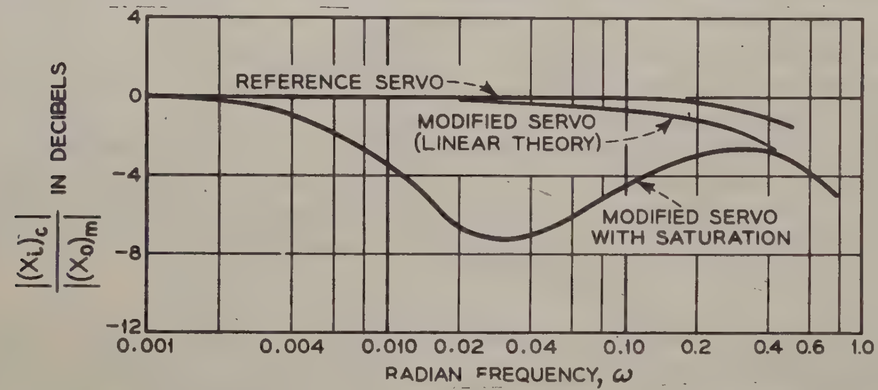


Fig. 12

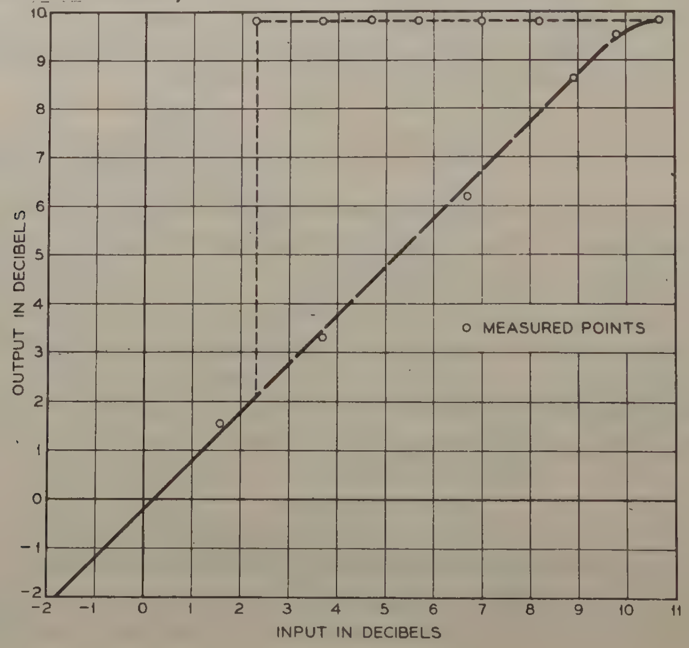


Fig. 13

valued, as the theory predicts. Furthermore, the amplitude range over which anomalous behavior exists was measured as approximately 7.8 db. This checks closely with the theoretical results which give 7.2 db.

<u>Note</u>: It was found necessary when checking these steady state conclusions on an analog computer, to make allowance for the fact that the saturated output displacement of the motor is a sawtooth rather than a sine wave at relatively low frequencies. This is because only the maximum velocity limitation is important at low frequencies. Under these conditions the feedback voltage is also a sawtooth. This makes it difficult to explore the steady state performance of the system in the immediate neighborhood of saturation with sinusoidal input signals, because the difference or error signal between a sinusoidal input signal and a sawtooth feedback signal of approximately equal amplitude has a large third harmonic term. As a result, under some conditions the error can change sign three times in a half cycle of the fundamental. To overcome this difficulty, a sawtooth rather than a sinusoidal input signal was used. It was not expected that this would substantially change the results since the amplitude of the third harmonic in a sawtooth is 19 db below the fundamental. This expectation is borne out by the fact that the results of the computations assuming sinusoidal inputs check very closely with the analog computer measurements using the sawtooth input signals.

Input-Error Performance

The steady state analysis shows that anomalous behavior in the input-output characteristics of a saturable system is accompanied by a double valued input-error characteristic. These results are more difficult to analyze because the error signal does not stay fixed under saturated conditions as does the output. This is shown by Eq. (15).

Therefore, for comparison purposes, we have computed the error to input signal ratio for the modified system at the critical input amplitude $|(X_i)_c|$, under the assumption that the output amplitude has jumped to its saturated value, and under linear conditions. The equation for the error under saturated conditions is given by:

$$\frac{|X_e|}{|(X_i)_c|} = \cot\theta. \qquad (19)$$

Figure 14 shows this ratio of $\frac{|X_e|}{|(X_i)_c|}$ for the modified servo system. Figure 14 also shows the linear case where

$$\frac{|X_e|}{|(X_i)_c|} = \frac{1}{1+G}.$$

The modified system becomes effectively linear at frequencies above 0.6 radians per second, where anomalous behavior ceases, and the linear equations were used to extend the analysis up to 2 radians per second. The reference servo gave results similar to the linear case in Fig. 14.

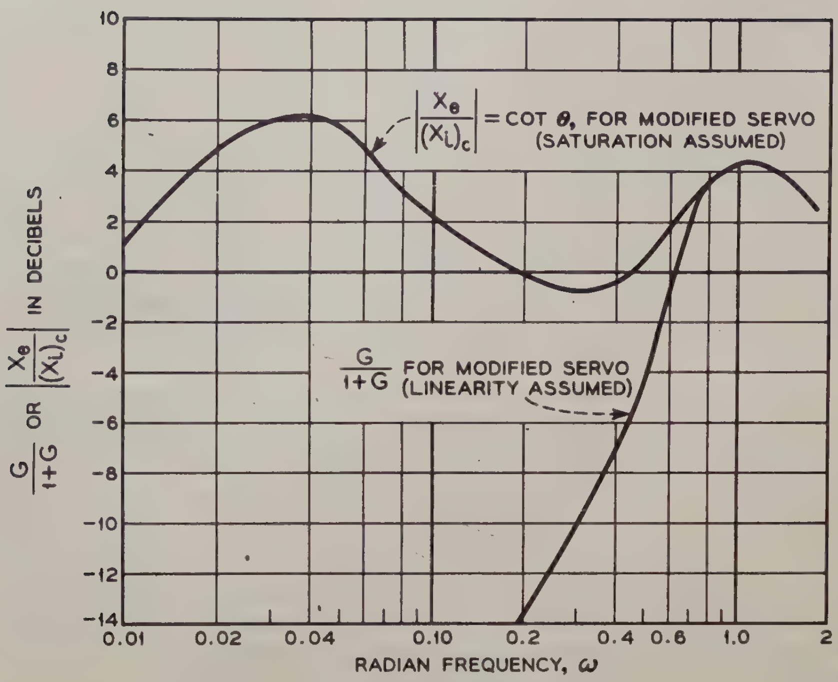


Fig. 14

As a comparison of the curves in Fig. 14 will show, the error to input ratio is greatly increased by anomalous behavior in the case of the modified servo, with a peak greater than unity occurring roughly at w = 0.03 radians per second. This corresponds to the dip in effective load carrying capacity that occurs at the same frequency, and for the same reason. The phase margin has a minimum at this frequency. The undesirability of an enhanced error characteristic of the type shown for the modified servo system is self evident.

Input-Error Measurements

Measurements of this hysteresis effect on the input vs. error characteristics were also made on the analog computer for the modified servo at w = 0.0316 radians per second. These results are shown in Fig. 15. The computed jump in error at the critical input amplitude is 46 db and the measured value is 45.5 db.

TRANSIENT CONSIDERATIONS

The large signal analysis presented above is restricted to single frequency signal conditions and so is not applicable to transient problems. The linear theory is not applicable either when the transients are large enough to produce saturation in the system. However, some useful information can be obtained from studying specific cases.

In a saturable servo system a large input step function will initially reduce the gain around the loop through compression in the limiter. The gain will recover in time as the output approaches the new equilibrium position required, and the fed-back signal restores the balance at the input. Hence, the gain is changing under these transient conditions. If the system is one

like the reference system analyzed above, in which the anomalous steady state behavior is negligible, the form of the transient recovery can generally be determined from a knowledge of the maximum velocity and the linear transient response to a step function just large enough to produce saturation.* When the system is like the modified system analyzed above, the anomalous steady state behavior is severe, and it is not likely that the transient could be represented by a simple damped exponential. However the transient recovery from a large step function is likely to be seriously degraded. Certainly the transient behavior would be less damped, and frequency of the oscillating component would be lower if the loop gain were artificially held fixed at a substantially reduced value.

Transient Measurements

The conclusion that a system exhibiting severe anomalous steady state behavior will also exhibit a degraded transient behavior in recovering from saturation is borne out by the transient response measurements made on an analog computer set up to simulate the reference and modified servo systems. Figure 16 shows the results for the two systems for a step function input.

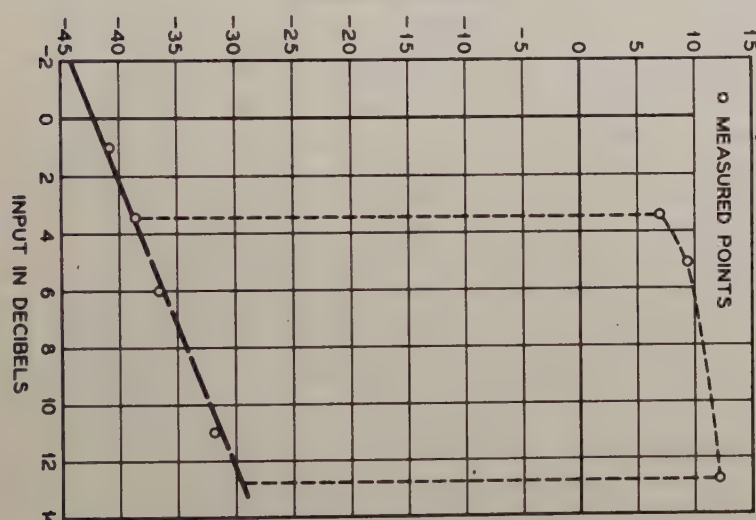


Fig. 15

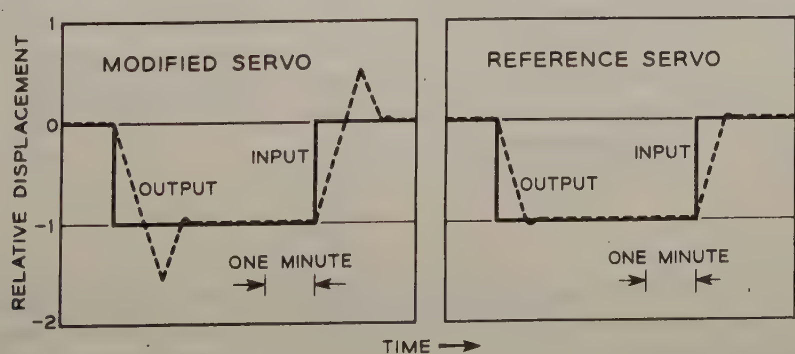


Fig. 16

*To satisfy this requirement for negligible anomalous steady state behavior for $|G|$ large, the minimum phase margin should be 60° $[|(X_i)_c| = 0.87\ |(X_o)_m|]$. Then the transient behavior at reduced gains would not be under damped.

The step input was 5 times as large as the level that would just saturate the system. The reference servo showed an overshoot of approximately 1/10 of the linear range of operation, and the transient died out is less then 6 seconds. The modified servo, on the other hand, had a transient overshoot 2 ½ times the linear range of operation, required 24 seconds to return to the balance point and then showed a second and third overshoot.

Stability vs. Noise

The phenomenon of a servo system becoming less stable in the presence of noise is readily explainable if the system is a saturable one in which the phase margin is substantially less at frequencies below gain cut-off. Then noise of sufficient amplitude to produce limiting would reduce the loop gain to desired signals and thus effectively reduce the phase margin of the system at gain crossover. Furthermore, if the noise were of much higher frequency than the signals, the amplitude of the noise required to produce velocity or acceleration limiting would be correspondingly less than that of signals. This follows from the fact that the load carrying capacity of the system falls off with frequency.

DESIGN IMPLICATIONS

The emphasis here is on understanding this anomalous behavior in saturable systems rather than on their design. However, it should be recognized that to eliminate anomalous behavior by restricting designs to ones with sufficiently large phase margins at high loop gains is not realistic. For other reasons it may not be desirable or even possible to do this. It is more realistic to treat these gain and phase indications of anomalous behavior as a warning of when and where these troubles will arise. Other techniques for minimizing this anomalous behavior can then be developed where necessary to fit particular cases. For example, in the modified servo analyzed above, it should be possible to greatly minimize the anomalous effects by the use of a limiter to restrict the amplitude of the voltage that can be stored on the capacitor in the integrating network.

CONCLUSION

The analysis shown here is built around the saturated transfer characteristics of the control loop. In some of the more complex saturated servo systems it is doubtful whether the necessary saturated transfer characteristics even exist as a single valued entity. However, in these systems where it does exist, the analysis shows that the dual mode of response exists in some degree at any frequency for which

$$|G| \cos \theta > 1 \cdot \qquad (9)$$

where $|G|$ is the magnitude of the gain and θ the phase margin of the saturated transfer characteristic of the loop. Furthermore, when this condition for anomalous behavior exists, the input-output characteristic will exhibit a dual response when

$$X_i > X_{o\ max} \sin \theta .$$

Here, X_i is the input signal, and $X_{o\ max}$ is the maximum output amplitude off the system at the frequency in question and θ is again the phase margin of the saturated loop transfer characteristic. For these inputs, the output will be either the linear value given by

$$X_o = X_i \frac{G}{1 + G}$$

or the saturated value $X_{o\ max}$.

The error will likewise be double valued. It will be either the linear value,

$$X_\varepsilon = X_i \frac{1}{1 + G}$$

or something between this linear value and $|X_\varepsilon| = |(X_i)_c| \cot \theta$, where $(X_i)_c$ is the critical input value at which dual response begines, as defined above. The precise value of $|X_\varepsilon|$ in this latter case depends on the value of $|X_i|$.

The predicted anomalous behavior for the two examples used for demonstration purposes is closely borne out by the measured results for these systems on an analog computer. Figures 12 and 13 show the input-output hysteresis, and Figs. 14 and 15 show the corresponding effects on the error.

REFERENCES

1. Tustin, A.: "Automatic and Manual Control." Academic Press (1952).

2. Levinson, E.: "Some saturation phenomena in servomechanisms with emphasis on tachometer stabilized systems." A.I.E.E. Technical Paper, pp. 53-110.

3. Tustin, A.: "A method of analyzing the effects of certain kinds on nonlinearities in closed cycle control systems." J.I.E.E. Vol. 94-1947 Part 11A, No. 1.

4. Kochenburger, R. J.: "A frequency method for analyzing and synthesizing contactor servomechanisms." Tr. A.I.E.E. 69, Part 1, pp. 270-284 (1950).

A NUMERICAL METHOD FOR DETERMINING A SYSTEM IMPULSE RESPONSE FROM THE TRANSIENT RESPONSE TO ARBITRARY INPUTS

Norman J. Zabusky
Hughes Aircraft Company
Los Angeles, California

Summary -- The transfer functions of large physical systems must be known if they are to be controlled efficiently. Environmental transient test procedures are described and their advantages presented. Because impulse and step functions cannot be obtained from practical test equipment, the problem of deducing the impulse response from arbitrary input and response data arises.

A successful iteration procedure for determining $h(\tau)$, the impulse response, by using the convolution equation, $r(t) = \int_0^\infty \theta(t-\tau)\, h(\tau)\, d\tau$, is described. An actual test transient is analyzed. An analog computer technique which mechanizes this iteration procedure is presented. The Appendix contains a tape multiplication procedure for performing numerical convolutions.

TESTING FOR DYNAMIC CHARACTERISTICS

Man always tests what he builds in order to determine if his design procedures are adequate and to see if his constructions meet desired specifications. In recent years dynamic testing of large physical systems, such as aircraft, has become increasingly important. One can design a better automatic pilot with an accurate knowledge of an aircraft's dynamic characteristics.

The best type of test is one which places the object in its natural environment, whether that be pressure, temperature or frequency. For example, if a high frequency filter is built with components whose values have been measured at dc or in the low frequency region, then large errors will arise in the over-all transfer function. Stray capacitances and inductances which do not show up in low frequency measurements cause these errors. An analogous situation is found with aircraft. At present they are tested in wind tunnels, and the information obtained allows a prediction of the transfer functions describing their actual "in-flight" characteristics. The wind tunnel is a simulated environment in that an artificial airstream creates forces and moments on a small model poised in various positions. As a final check on the "in-flight" dynamic characteristics, it is desirable to use the natural environment of "free-flight tests."

Before testing a system, one should have a rough idea of its characteristics, such as orders of magnitude to expect in recorded quantities, the order of the differential equations describing the system and what nonlinearities to expect. This will allow a more effective and meaningful test, and one to which data reduction techniques are most amenable. Care must be exercised when testing nonstationary systems. For aircraft this means a constant velocity and altitude during each test, as well as small perturbations around the trim or quiescent condition.

TEST PROCEDURES

Dynamic test procedures are broken down into two essential classes: steady-state and transient. The first procedure is familiar to most engineers and is one in which the input to the system is excited sinusoidally, and the response is measured at different frequencies. This procedure requires exceptional test and measuring equipment. It is very hard (and sometimes impossible) to get good sine waves from the elevators of aircraft. Furthermore, steady state tests take a long time and require an object which is stationary with respect to its environment. Because of the length of the test, the stationary requirement is sometimes impossible to meet. These difficulties are the reasons for transient testing.

In transient tests one excites the input (elevator motion) with some function of time and records the response. This gives one a maximum of information in a minimum of time and recording space. For example, the damping factor, ξ, of an underdamped second order system can easily be determined by exciting it with a step function. Figure 1a gives ξ vs. R, where R is determined from the quotient of "amplitude-deviations" of the step function response, as shown in the formulas in Fig. 1b. I have used this procedure for checking the damping of aircraft simulated on analog computers. It gives an accuracy of from 1% to 3%.

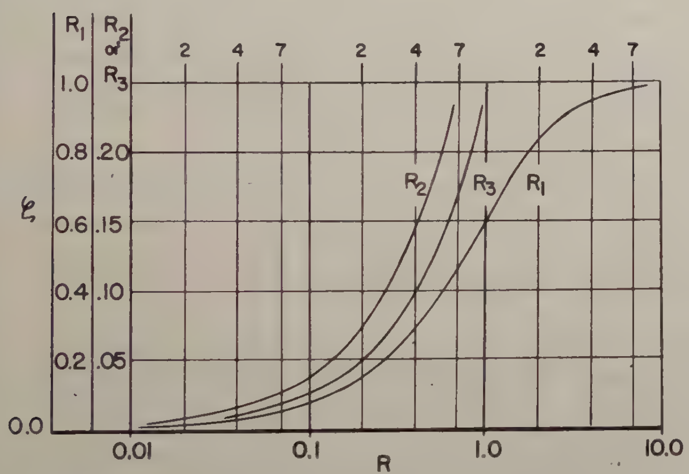


Fig. 1a - ξ vs. decrement ratio for a second order system.

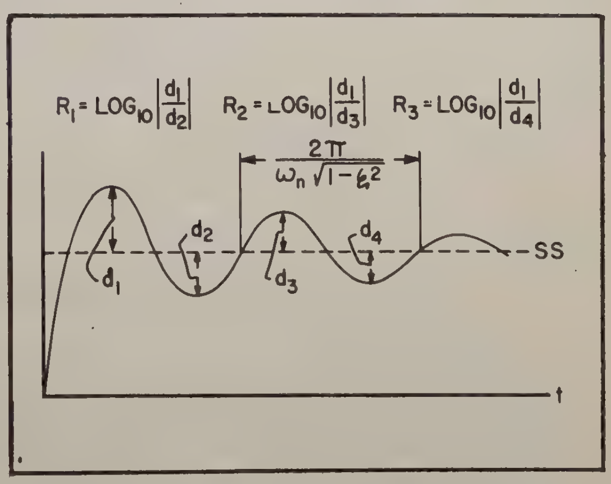


Fig. 1b - The step response of a second order system.

Most often, idealized inputs as steps and impulses cannot be obtained in practice. Therefore, we are forced to use what we can get from our programming equipment. The result is usually an arbitrary function (Fig. 2) which cannot be conveniently described mathematically. One chooses the general shape of the arbitrary input so that it reveals characteristics that one is most interested in seeing: to avoid nonlinearities, small signals are used; to bring out high frequency effects, rapidly varying input functions are used.

It is most desirable to excite the system with pulse-type and doublet-type inputs. An example of the latter is shown in Fig. 2. These minimize the

disturbances to the system from trim condition and contain a maximum of information about the natural behavior of the system. Step-type inputs frequently hide the natural behavior in the steady-state value reached. Sometimes it is necessary to test the system with several inputs. To cover the frequency spectrum of interest the duration of the input function is varied, and to determine the linear range the amplitude of the input function is varied. This will require adjusting sensitivities and ranges of instruments to each input.

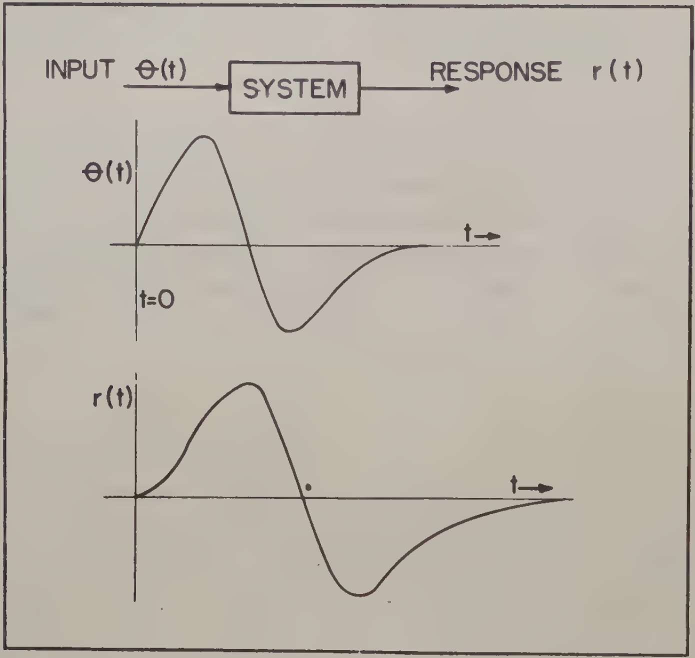


Fig. 2 - Typical input and response curves.

The lack of an analytical input poses the problem to be solved, namely: deduce the impulse response of a system when the input and response are graphically recorded functions of finite accuracy. Many sources of inaccuracy are present during testing, some of which are shown in Fig. 3. In most practical applications one can rarely expect to know the full scale quantity being measured to more than three significant figures. Usually 3% of full scale is considered good. These errors must be accounted for in order to determine a reasonably good approximation to the true impulse response.

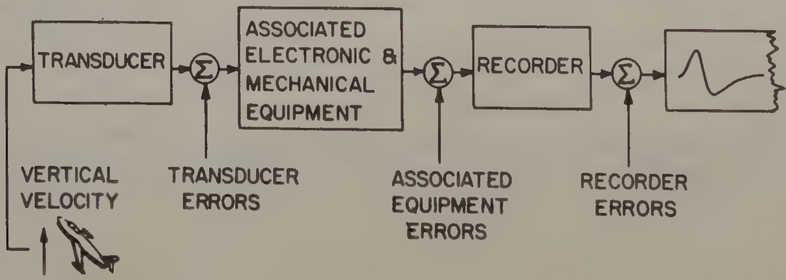


Fig. 3 - A sequence for recording a physical quantity.

BRIEF HISTORY AND COMMENTS ON NUMERICAL PROCESSES

Many investigators (see References) have developed techniques for determining the dynamic characteristics of systems from test data. In many cases the techniques are for a specific system or a specific form of test data. Most deal with linear systems. However, Shimbrot treats systems with nonlinearities whose shape is known a priori. Note that a great deal of the work has been done by people in the field of aerodynamics. Milliken's paper presents a thorough summary of the work done in the United States (previous to 1951) on determining transfer functions from flight-test data. He includes a very comprehensive bibliography.

Almost all the techniques listed in the bibliography are numerical. Most of them deal with the test data in the time domain, using the differential equations of motion of the system as a basis for analysis. In the following section another technique for determining the impulse response in the time domain will be described and illustrated. It assumes linearity and is based on the convolution equation.*

Numerical analysis is applied in almost all the above techniques, whether it is done consciously or unconsciously. One must be very careful with these applications, otherwise erroneous answers will result. Many problems can be formulated and solved analytically. However, the transferal of the analytic methods to the domain of numerical methods can sometimes lead to false solutions or difficulties in obtaining the true solutions.†

The answer (impulse response) obtained by a particular numerical technique is a function of the technique and the data. For the problem posed above, there is no unique answer.

CONVOLUTION EQUATION TECHNIQUE FOR DETERMINING THE IMPULSE RESPONSE

Approximating the Impulse Response

This technique makes use of the convolution equation‡

---

*Zabusky, N. J.: "Numerical Methods for Determinging Impulse Responses from Recorded Data." Master's thesis in electrical engineering, M.I.T., 1951. This paper has also been published by the Servomechanisms Laboratory of M.I.T.

†A classic example of this arises when solving differential equations by difference techniques. If the basic spacing chosen for the difference approximation is too large one can get false answers. For example, when solving the differential equation of an undamped second order system (linear oscillator), too large a spacing will cause the introduction of damping and a decrease in the natural frequency of oscillation. See the Houbolt approximation in the following book: Levy, S. and Kroll, W.: "Errors Introduced by Finite Space and Time Increments in Dynamic Response Calculations," Proceedings of the First National Congress of Applied Mechanics, pp. 1-9 (1951).

‡Gardner, M. F., and Barnes, J. L.: "Transients in Linear Systems." New York, John Wiley and Sons, Inc., pp. 231 and 262 (1950).

$$r(t) = \int_0^{\infty} \theta\,(t - \tau)\; h_t\,(\tau)\; d\tau \tag{1}$$

where:

$\theta(t)$ is the true (or given) input function,
$r(t)$ is the true (or given) response function,
$h_t(\tau)$ is the true impulse response of the system.

The problem of determining the impulse response from the given input and response data is one of solving an integral equation. The convolution equation is a Volterra Integral Equation of the First Kind. To accomplish this solution the true impulse response $h_t(\tau)$ is approximated with the function $h(\tau)$.

$$h(\tau) = x_1\, h_1\,(\tau) + x_2\, h_2\,(\tau) + \ldots.\; x_m\, h_m\,(\tau) \tag{2}$$

In Eq. (2) it is assumed that the functions $h_i(\tau)$ are known. The x_i are the unknown constants which characterize the impulse response and are determined to "best" satisfy the convolution equation. The choice of the functions $h_i(\tau)$ is motivated by physical considerations. The impulse response of any system described by linear differential equations with constant coefficients can be a sum of polynomials, exponentials, sinusoids and/or any combination of the products of these three functions. Hence the <u>basic approximating function</u> is the exponential-polynomial product, or

$$h_i(\tau) = e^{x_{n+1}\tau}\,(x_1 + x_2\tau + x_3\tau^2 + \ldots.\; x_n\tau^{n-1}). \tag{3}$$

In the most general case, the exponent x_{n+1} is a complex number.

Solution of the Problem

The technique for determining the unknown coefficients, x_i, is numerical. The basic numerical process used is that of "performing a convolution." This is the process of evaluating the integral of Eq. (1) when θ and h are known. A "tape multiplication" scheme for accomplishing this is described in the Appendix.

The technique for determining the impulse response employs an iteration process. The exponents are modified from one iteration to the next, and at each step the coefficients of the exponential-polynomial products are determined. With a knowledge of these coefficients, the exponents are modified and the iteration process repeated until a "best" solution is obtained. The number [subscript m in Eq. (2)] of basic approximating functions that should be used is equal to the order of the system being analyzed. Frequently, the physics of the situation will give the required order; if not, one can guess at the order. After a few iterations he will determine whether this guess yields an accurate fit or whether an additional basic approximating function must be added to $h(\tau)$. These points are illustrated by the following example:

Suppose $r(t)$ and $\theta(t)$ are given as graphically recorded functions. Assume that $h(\tau)$ has the form,

$$h(\tau) = x_{10}\, e^{x_{20}\tau} + x_{30}\tau\, e^{x_{20}\tau} = e^{x_{20}\tau}\,(x_{10} + x_{30}\tau). \tag{4}$$

This second order impulse response is used when one suspects that the unknown system is of <u>first order</u>. <u>The additional order provides the flexibility required to iterate</u>. In order to construct the h tapes for convolution, x_{20} must be known. Usually some a priori information on the system's behavior will furnish a value of x_{20}. If no information is available, a procedure which decomposes* the given transients and yields a crude step-response is recommended. Otherwise, one must make a wild guess as to the value of x_{20} from the shape of the given transients. The zero in each subscript of x represents the first approximation for each coefficient. The next step is to solve for x_{10} and x_{30}. Substitute Eq. (4) into Eq. (1).

$$r_c(t_i) = x_{10}\int_0^{t_i} \theta(t_i - \tau)\, e^{x_{20}\tau}\, d\tau + x_{30}\int_0^{t_i} \theta(t_i - \tau)\, \tau e^{x_{20}\tau}\, d\tau. \tag{5}$$

The subscript c on $r_c(t_i)$ designates the "computed" response. This is to distinguish it from the "given" response $r(t_i)$. The upper limit ∞ can be replaced by t_i if $\theta(t) = 0$ for $t < 0$. If the need arises, this can be accomplished by a translation of coordinates.

If we let

$$a_{i1} = \int_0^{t_i} \theta(t_i - \tau)\, e^{x_{20}\tau}\, d\tau \tag{6}$$

$$a_{i3} = \int_0^{t_i} \theta(t_i - \tau)\, \tau\, e^{x_{20}\tau}\, d\tau \tag{7}$$

then

$$r_c(t_i) = x_{10}a_{i1} + x_{30}a_{i3}. \tag{8}$$

The "convolution coefficients" a_{i1} and a_{i3} are evaluated numerically as described in the Appendix. If we let $r_c(t_i) = r(t_i)$, we can obtain as many equations as there are values of $r(t_i)$ given. For example,

$$\begin{aligned}
r(t_1) &= x_{10}\, a_{11} + x_{30}\, a_{13} \\
r(t_2) &= x_{10}\, a_{21} + x_{30}\, a_{23} \\
r(t_3) &= x_{10}\, a_{31} + x_{30}\, a_{33} \\
&- - - - - - - - - - - - - \\
r(t_n) &= x_{10}\, a_{n1} + x_{30}\, a_{n3}
\end{aligned} \tag{9}$$

*Zabusky, <u>op. cit.</u>, pp. 127-136.

In matrix form this is

$$r(t_i)\Big] = \Big[a_{ij}\Big] \Big[x_{jo} = \Big[a_{ij}\Big] \begin{bmatrix} x_{10} \\ x_{30} \end{bmatrix} \qquad (10)$$

Now x_{10} and x_{30} are to be determined so that $r_c(t_i)$ will agree most closely with the given $r(t_i)$. Two procedures are available. In the first we solve for x_{10} and x_{30} using all the equations in (9) in a least square manner. This procedure minimizes the mean-square error ε_{MS},

$$\varepsilon_{MS} = \int_0^T \Big[r(t) - r_c(t)\Big]^2 \, dt. \qquad (11)$$

Equation (10) is substituted into Eq. (11) and the integration from $0 \rightarrow T$ is replaced by a finite summation. If the equations are manipulated, one obtains two linear simultaneous equations.

$$\Big[a_{ji}\Big] \Big[D\Big] \Big[r(t_i) = \Big[a_{ji}\Big] \Big[D\Big] \Big[a_{ij}\Big] \begin{bmatrix} x_{10} \\ x_{30} \end{bmatrix} \qquad (12)$$

<u>Note</u>: $j = 1$ and 3; $\boxed{a_{ij}}$ is an $i \times j$ matrix whose elements are the convolution coefficients and $\boxed{D}$ is a diagonal matrix whose elements are area-weighting factors. (These correspond to the integration from 0 - T. For Simpson's Rule these diagonal elements are 1, 4, 2, 4 ..... 4, 1.)

A much easier procedure fits $r(t)$ at as many points as there are unknown x's (two in this case). Hence $r_c(t)$ and $r(t)$ will agree exactly at two points.

I chose the second procedure because the first involves at least ten times as much numerical work for a fair sampling, and because erroneous solutions may arise when solving the simultaneous equations of (12). These false solutions result from the accumulation of truncation and roundoff errors in the additional integration and the matrix multiplications. The chief disadvantage of the second procedure lies in the fact that x_{10} and x_{30} are not only functions of the assumed x_{20}, but also of the values of t at which the curves were fitted. The first procedure is worthy of more investigation if accurate high-speed computing equipment is available.

When plotted, $r_c(t)$ will most likely differ from $r(t)$, except at the points at which they were fitted. If the physical system is truly of first order and if the effects of truncation and roundoff errors are negligible, the difference is due to a wrong value of x_2. To be able to iterate, a new value of $x_2 = x_{21}$ must be obtained.

$$x_{21} = x_{20} + \frac{x_{30}}{x_{10}} (k)^{*} \tag{13}$$

where k is a convergence factor. In the examples chosen a value of $k = 1.3$ was used. When starting the iteration procedure, values of k such that $0.80 \leq k \leq 1.30$ are acceptable. After a few iterations are performed, k should always be chosen less than 0.80. The choice of k will determine whether the iteration procedure is stable. With this new value of $x_2 = x_{21}$, the above process can be repeated. The iteration is continued until $x_3 \to 0$. If x_3 does not approach zero, one must consider four possibilities:

1. The system is of second or higher order.

2. The system has a double order pole at x_{20} (a rarity).

3. Inaccuracies in the original data and/or numerical truncation and round-off errors have introduced extraneous solutions.

4. The points at which $r_c(t)$ was fitted to $r(t)$ have introduced extraneous solutions.

In case 2, $r_c(t)$ will agree very closely with $r(t)$ and the problem can be considered solved. In cases 1, 3 and 4, $r_c(t)$ will differ appreciably from $r(t)$, and this leads one to use an impulse response of a fourth-order system.

$$h(\tau) = x_{10}\, e^{x_{20}\tau} \left(1 + \frac{x_{30}\tau}{x_{10}}\right) + x_{50}\, e^{x_{40}\tau} \left(1 + \frac{x_{70}\tau}{x_{50}}\right) \tag{14}$$

Actually, a second order system is suspected. The extra two orders in the impulse response gives one the flexibility for iteration, i.e., changing the exponents. After a number of iterations, one will obtain a better agreement between $r_c(t)$ and $r(t)$ than was obtained with the second order impulse response, Eq. (4). This process of introducing additional orders can be continued until the desired agreement is obtained between $r_c(t)$ and $r(t)$. Unfortunately, no straightforward methods are available for determining when extraneous solutions have been introduced. That is why one should have an a priori acquaintance with the system and its order. With this knowledge one

---

*This relation is obtained by letting $x_2 = x_{20} + \delta$ in the expression for $e^{x_2\tau}$.

$$e^{x_2\tau} = e^{x_{20}\tau}\, e^{\delta\tau} = e^{x_{20}\tau} \left(1 + \delta\tau \quad \frac{(\delta\tau)^2}{2} + \ldots\right)$$

If δ is very small, then, to a rough approximation, $\frac{(\delta\tau)^2}{2}$ and all remaining terms are negligible. By comparing Eq. (4) (whose first term is normalized) with the above equation, one obtains $\delta = \frac{x_{30}}{x_{10}}$. The convergence factor, k, takes into consideration the neglected terms in the series expansion.

can start with a good approximation of the impulse response and keep the order the same during the iteration. If erroneous solutions are suspected, one should decrease the basic spacing used in performing a convolution and repeat the procedure.

To aid the convergence of the iteration process, one can triple the order of the approximating impulse response. For example, if one suspects that the given data describes a first order system, he can use Eq. (15) rather than Eq. (4) to approximate $h_t(\tau)$.

$$h(\tau) = x_{10}\, e^{x_{20}\tau} \left(1 + \frac{x_{30}}{x_{10}}\tau + \frac{x_{40}}{x_{10}}\tau^2\right). \tag{15}$$

The values of $\frac{x_{30}}{x_{10}}$ and $\frac{x_{40}}{x_{10}}$, when found, can be used to give a more accurate value of δ than that obtained from Eq. (13). This procedure increases the work involved per iteration but will decrease the number of iterations required to obtain a good solution.

The procedures described above can be adapted to systems whose poles are complex-conjugate. An iteration procedure based on the minimization of a defined error quantity, as ϵ_{MS}, will probably prove most advantageous.

Example Problem

Figure 4 shows the input and response of an actual physical system. From a priori information we write $h(\tau)$ as

$$h_0(\tau) = x_{10} e^{-1.40\tau})1 + \frac{x_{30}}{x_{10}}\tau) + x_{50} e^{-7.35\tau} (1 + \frac{x_{70}}{x_{50}}\tau) \tag{16}$$

where

$$x_{20} = -1.40$$
$$x_{60} = -7.35. \tag{17}$$

The input was convolved numerically with Eq. (16), and the result fitted at the values of t aligned with the black arrows (trial 0). Although there are only four unknowns in Eq. (16), five equations, and hence five unknowns, are indicated by the five black arrows. The extra unknown is a bias which is added to $r_c(t)$. This takes into consideration unknown biases which arise when reading the recorded data. Hence, the general equation used in fitting will be of the form

$$r(t_i) = x_{10}\, a_{i1} + x_{30}\, a_{i3} + x_{50}\, a_{i5} + x_{70}\, a_{i7} + x_{80}. \tag{18}$$

where x_{80} is the bias. The trial 0 impulse response takes the form

$$h_0(\tau) = 2.66\ e^{-1.40\tau} (1 - .764\tau) + 4.33\ e^{-7.35\tau} (1 - 2.73\tau). \tag{19}$$

The bias $x_{80} = +4.8$.

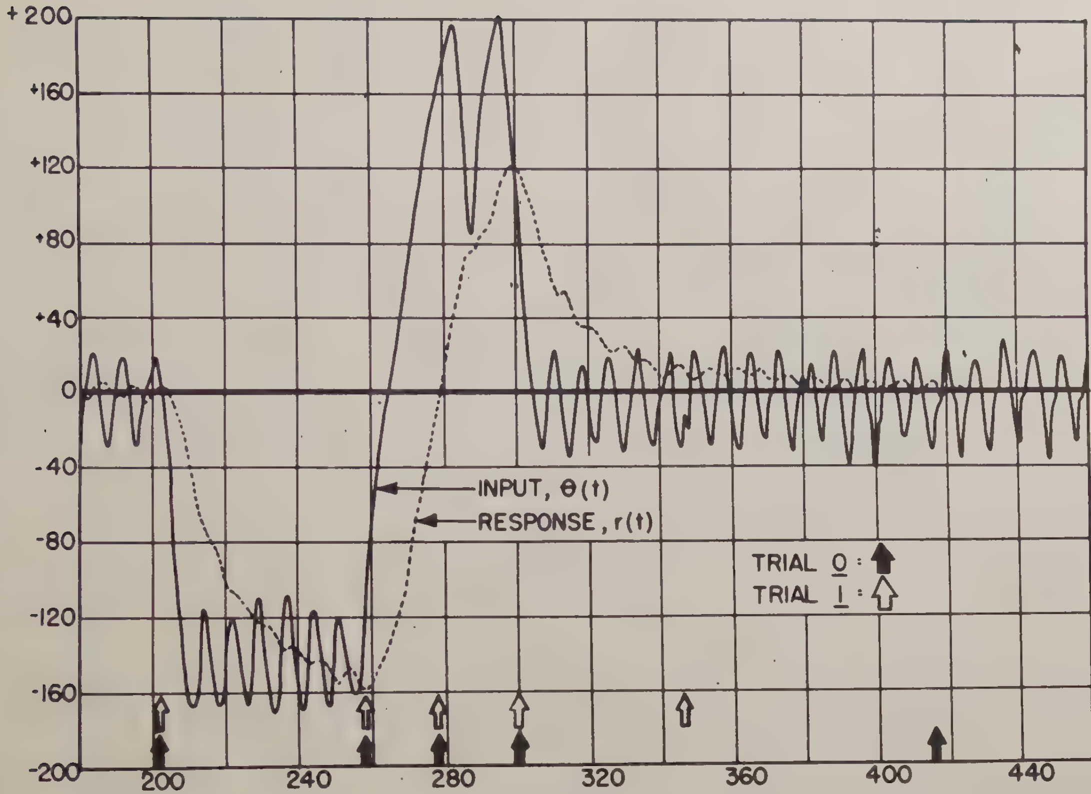


Fig. 4 - The input and response functions obtained from oscillograph film.

From this, δ is computed as (k = 1.30)

$$\delta_{21} = 1.3\ (-.764) = -.990$$

$$\delta_{61} = 1.3\ (-2.73) = -3.55 \tag{20}$$

Hence $x_{21} = -2.39$ and $x_{61} = -10.9$. The impulse response for the trial 1 iteration becomes

$$h_1(\Upsilon) = x_{11} e^{-2.39\Upsilon}\left(1 + \frac{x_{31}}{x_{11}}\Upsilon\right) + x_{51} e^{-10.9\Upsilon}\left(1 + \frac{x_{71}}{x_{51}}\Upsilon\right) \tag{21}$$

The five white arrows indicate the points of fitting for trial 1. The following was determined as $h_1(\Upsilon)$:

$$h_1(\Upsilon) = 1.48\ e^{-2.39\Upsilon}(1 + .358\Upsilon) + 4.76\ e^{-10.9\Upsilon}(1 + 6.40\Upsilon). \tag{22}$$

The bias for trial 1 $x_{81} = +6.05$. Figure 5 shows the error between r(t) and $r_c(t)$ as a percentage of full scale, that is

$$\epsilon(t) = \frac{100\%}{280}\left[r(t) - r_c(t) + x_8\right]. \tag{23}$$

Note the appreciable reduction in the error obtained by just one iteration.

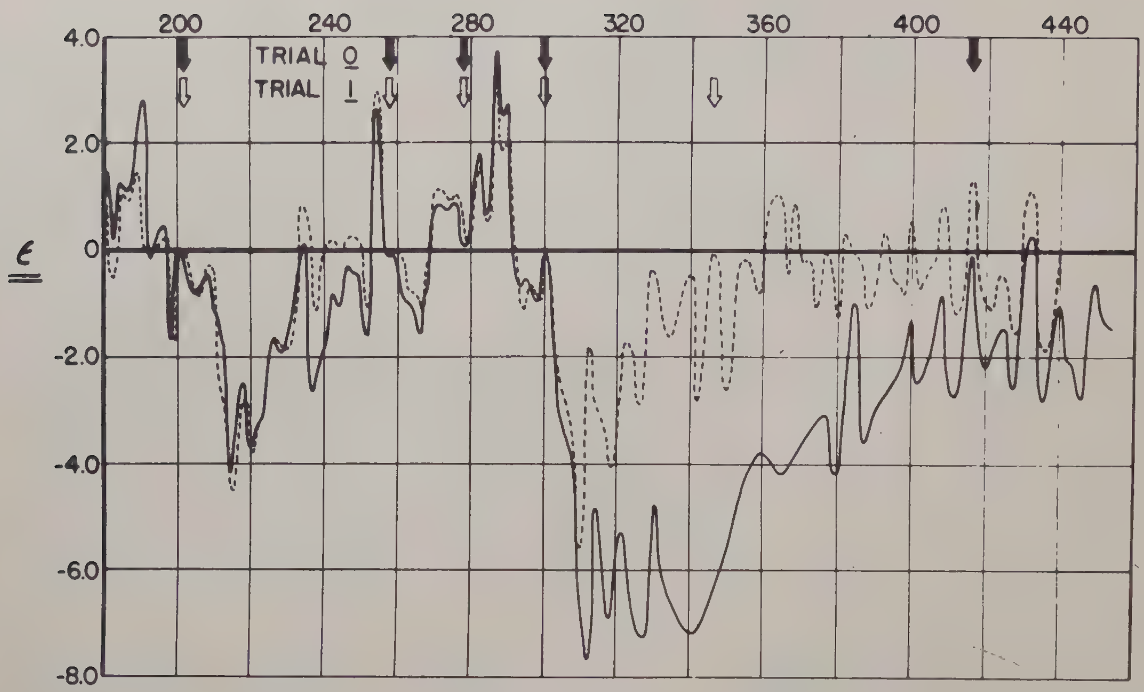


Fig. 5 - Per cent error as a function of t.

ANALOG COMPUTER TECHNIQUE FOR DETERMINING THE IMPULSE RESPONSE

The above procedures are all numerical and as such are adaptable to solution by means of desk calculators or digital computers. I would like to suggest one of possibly many analog computer techniques that can be used to solve this problem.

Figure 6 shows how this technique is mechanized. The given input $\theta(t)$ and the given response $r(t)$ are produced by analog function generators. The input is fed to the analog computer and the response of the computer, $r_c(t)$ is mixed with the given response $r(t)$ and a proper bias. The result is fed to the circuits shown. The output of the top integrator will then be the mean-square error, MS.

$$\mathcal{E}_{MS} = \int_0^T \left[r(t) - r_c(t) + (\text{bias}) \right]^2 dt \tag{24}$$

and the output of the lower integrator will be the mean-absolute-value error, $\mathcal{E}_{MA}$

$$\mathcal{E}_{MA} = \int_0^T \left[\left| r(t) - r_c(t) + (\text{bias}) \right| \right] dt. \tag{25}$$

The correct impulse response is obtained when both these quantities are absolute minima. For example, if x_3, x_4 and the bias are all zero, then the error, $\mathcal{E}$, will be only a function of x_1 and x_2. This can be depicted by the surface shown in Fig. 7. By suitable adjusting potentiometers x_1 and x_2, the point $(x_1)_{min}$ and $(x_2)_{min}$ will be reached. These two quantities then charac-

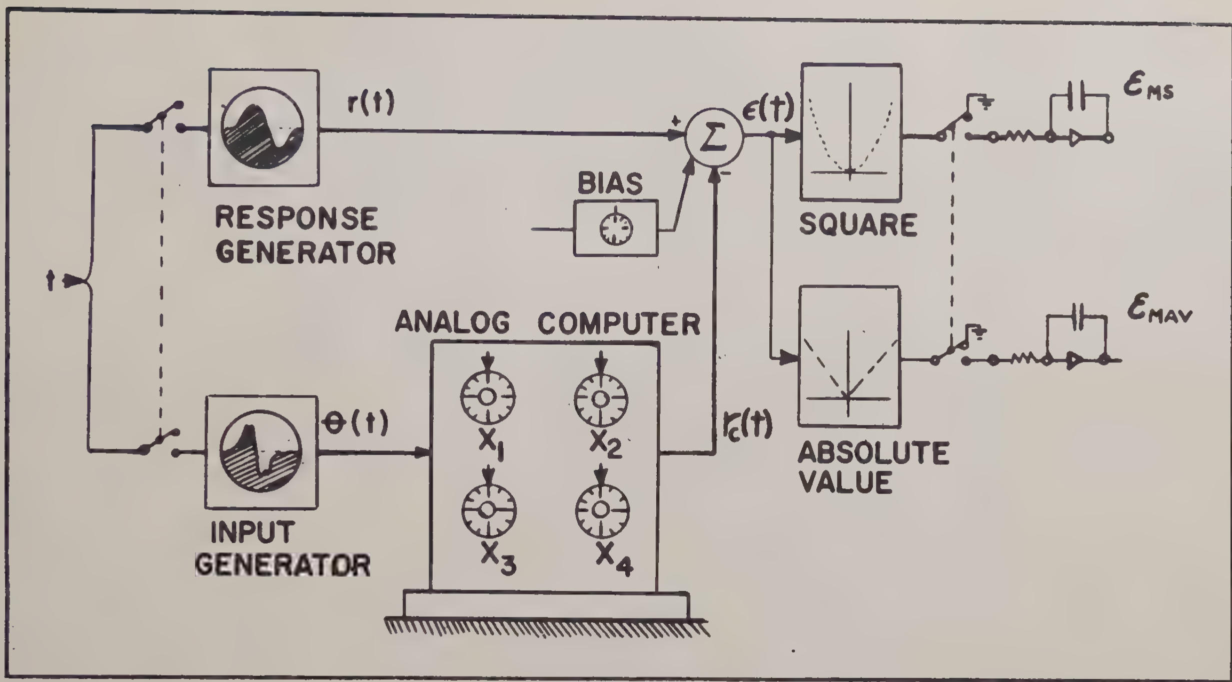


Fig. 6 - Determination of the impulse response using an analog computer.

terize the correct impulse response. The logical sequence for varying these potentiometers has been described mathematically.* In essence, one seeks the point in the error space at which the first partial derivatives of the error are all zero, that is

$$\frac{\delta \epsilon}{\delta x_i} = 0 \quad , \qquad i = 1, 2, 3, 4. \tag{26}$$

This is a necessary condition.

CONCLUSION

A numerical iteration procedure for determining the impulse response of an overdamped system has been demonstrated. The error of a successful iteration procedure must converge to a volume in error space which satisfies given requirements, and the rate of this convergence must be rapid. Since convergence depends on the "general shape" and accuracy of the data, no positive statements can be made. Future work on this problem should be directed along these lines.

It is to be noted that the original data was used in all the above calculations. To treat "noisy" data conveniently one must calculate the input auto-correlation function, $\varphi_{ii}(t)$, and the input-response cross-correlation function, $\varphi_{ir}(t)$ from the given data. These are related by the same integral equation:

$$\varphi_{ir}(t) = \int_0^{\infty} h_t(\tau)\ \varphi_{ii}(t - \tau)\ d\tau$$

*Zabusky, op. cit., pp. 147-176.

ACKNOWLEDGEMENT

I wish to thank Professor G. C. Newton, Jr., of the Electrical Engineering Department at the Massachusetts Institute of Technology for his advice and guidance during this research investigation. Most of the work was performed at the Servomechanisms Laboratory, M.I.T., and was sponsored by the Bureau of Ordnance, U.S. Navy, under Contract NOrd-11799.

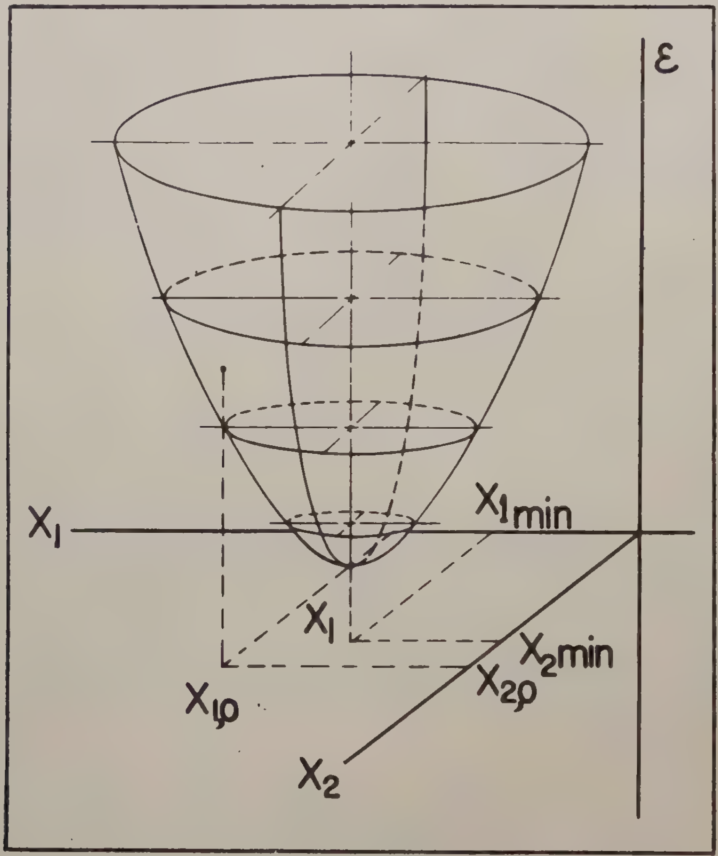


Fig. 7 - The error in three dimensional space.

APPENDIX: PERFORMING A CONVOLUTION

To visualize the process of convolution, first examine the nature of the function, $\theta(t - \tau)$, designated as the shifting input. The shifting nature can be visualized by considering the simple example in Fig. 8 (A, B and C). Here $\theta(\tau) = + \tau$ is a straight line of slope + 1. $\theta(- \tau)$ is obtained by replacing $+ \tau$ by $- \tau$, and the result is a straight line of slope - 1. $\theta(- \tau)$ is a mirror image of $\theta(\tau)$ around $\tau = 0$. If we replace $(- \tau)$ by $(t - \tau)$, then $\theta(t - \tau) = t - \tau$ is a straight line of slope - 1, shifted in the positive direction by an amount $\tau = t$. A completely analogous discussion holds for any arbitrary function as shown in Fig. 8 (D, E and F). As t takes on increasing values, the mirror image (or folded) curve shifts along the abcissa in the positive τ direction.

To perform a convolution, one shifts the input function by an amount t, multiplies its ordinates with corresponding ordinates of the impulse response

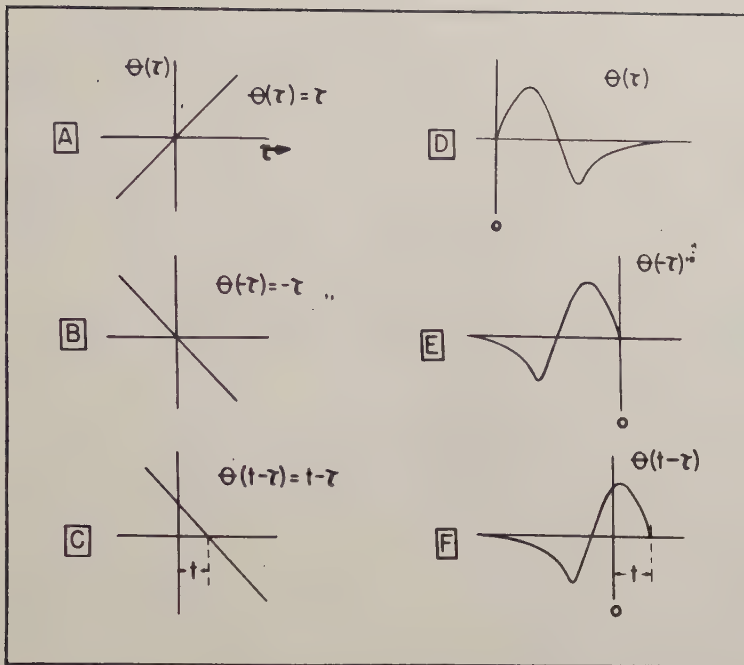


Fig. 8 - The shifting input.

and obtains a product curve, $P(t - \tau)$, as shown in Fig. 9. The area under the product curve is the required convolution.

Numerical convolution is accomplished easily with convolution tapes.* They systematize the calculations for desk calculating machines and minimize the writing and thought required. The "h" tape is constructed by recording in boxes the ordinates of the h curve at an equal spacing Δ. (See Fig. 10.) This tape starts at $\tau = 0$ and runs toward the right, just as the impulse response starts at $\tau = 0$ and runs toward the right. The θ tape starts at zero and runs toward the left. In each box of this tape one records an ordinate of $\theta(\tau)$ each weighted by a factor depending upon the numerical integration scheme being used. Simpson's Rule is used in the θ tape of Fig. 10.

To evaluate $r_c(8\Delta)$, one shifts the θ tape to the right by 8Δ units as shown in Fig. 10. Then one multiplies the values of the ordinate h by the weighted value of θ in the box immediately below, and adds the product to the previous products. The quantities in the second row of the θ tape are used only for the last multiplication in any convolution. Hence

$$r_c(8\Delta) = \frac{\Delta}{3}\Big[\theta(0)\,h(8\Delta) + 4\,\theta(\Delta)\,h(7\Delta) + 2\,\theta(2\Delta)\,h(6\Delta) + 4\,\theta(3\Delta)\,h(5\Delta) + 2\,\theta(4\Delta)\,h(4\Delta) + 4\,\theta(5\Delta)\,h(3\Delta) + 2\,\theta(6\Delta)\,h(2\Delta) + 4\,\theta(7\Delta)\,h(\Delta) + \theta(8\Delta)\,h(0)\Big]. \qquad (27)$$

The operations of multiplication and accumulation are accomplished conveniently with a desk calculator.

---

*Zabusky, op. cit., pp. 78-87.

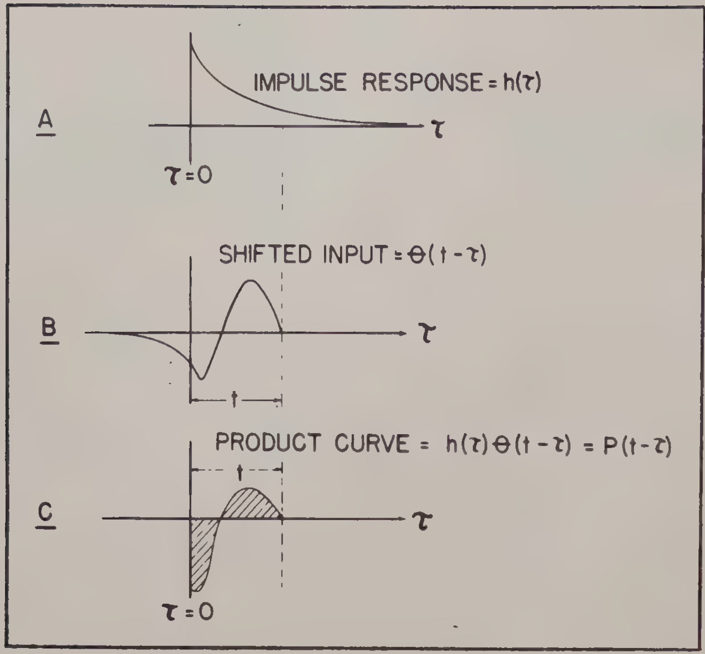


Fig. 9 - Visualizing a convolution.

The size of the interval Δ is determined by the shape of the product curve, $P\ (t - \tau)$. Δ should be small enough so that a parabolic approximation to any segment of the product curve of 2 Δ units in length will be an accurate approximation.

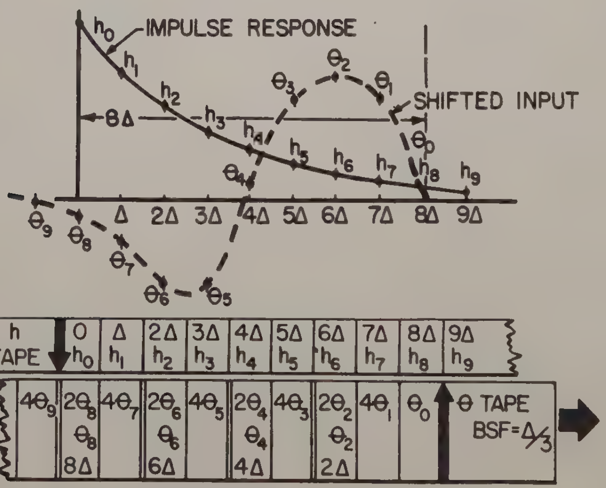


Fig. 10 - Convolution by tape multiplication.

REFERENCES

1. Donegan, J. J., Robinson, S. W., and Gates, O., Jr.: "Determination of Lateral Stability Derivatives and Transfer Function Coefficients from Frequency Response Data for Lateral Motions." National Advisory Committee for Aeronautics (NACA), T.N. 3083 (1954).

2. Donegan, J. J., and Pearson, H. A.: "Matrix Method of Determining the Longitudinal Coefficients and Frequency Response of an Aircraft from Transient Flight Data." NACA, Report 1070 (1952).

3. Eggleston, J. M., and Mathews, C. W.: "Application of Several Methods for Determining Transfer Functions and Frequency Response of Aircraft from Flight Data." NACA, T.N. 2997 (Sept. 1953).

4. Greenberg, H.: "A Survey of Methods for Determining Stability Parameters of an Airplane from Dynamic Flight Measurements." NACA, T.R. 2340 (Apr. 1951).

5. Madwed, A.: "Number Series Method for Solving Linear and Non-linear Differential Equations." Proc. IRE PGCT, of the Symposium of Non-linear Circuit Analysis, New York, p. 320 (Apr. 1953).

6. Milliken, W.: "Dynamic Stability and Control Research." Published by the Royal Areonautical Society, Third Anglo-American Aeronautical Conference (1951).

7. Seamans, R. C., Jr., Blasingame, B. P., and Clementson, G. C.: "The pulse method for the determination of aircraft dynamic performance." Jl. of Aeronautic Sci., vol. 17, no. 1 (Jan. 1950).

8. Seckel, E.: "A Method of Finding the Response of a Linear Dynamic System to an Ideal Step Function Input Given the Observed Response to an Arbitrary Observed Input." Cornell Aero. Lab., Flight Research Dept., Memo No. 66 (May 4, 1949).

9. Shinbrot, M.: "On the Analysis of Linear and Non-linear Dynamical Systems from Transient Response Data." NACA, T.N. 3288 (Dec. 1954).

10. Smith, A. G., and Triplett, W. C.: "Experimental Flight Methods for Evaluating Frequency Response Characteristics of Aircraft." Paper No. 54, SA-3, A.S.M.E. (1954).

11. Strumpers, F.: "A First Time Passage Problem." Philips Research Reports, R-144, pp. 270-281 (Aug. 1950).

12. Tustin, A.: "A method of analyzing the behavior of linear systems in terms of time series." J.I.E.E. (London) vol. XCIV, pt. II-A, no. 1 (1947).

SYNTHESIS OF FEEDBACK CONTROL SYSTEMS WITH A MINIMUM LEAD FOR A SPECIFIED PERFORMANCE*

George S. Axelby
Westinghouse Electric Corporation
Baltimore, Maryland

Summary -- Compensation is usually needed in a high performance feedback control system to realize a desired frequency response. Direct synthesis procedures to determine the necessary compensation are described. In particular, consideration is given to "minimum lead" systems which have an optimum gain.

The desired performance of a feedback control system may be specified approximately in different ways. In some applications, the transient response of a system to a particular input may be prescribed, and in other systems the frequency response may be specified.

Of course, these criteria are related to each other in a given system, but system performance may be conveniently specified approximately in terms of the frequency response, as summarized in Fig. 1. The open loop and closed loop frequency responses and transient responses of the system are indicated, as well as the approximate relations between them which hold for a large class of systems.

Using these relationships, the band-pass, ω_c, the position constant, velocity constant, the minimum phase margin or the phase margin, the frequency of minimum phase margin and the gain at particular frequencies may be determined approximately from desired performance specifications. Of course, the requirements may conflict if all variables are specified. In this case, the usual design compromises are necessary.

In the following discussions, it will be assumed that the desired system performance can be translated into the frequency response criteria shown in Fig. 1, and synthesis techniques will be established in terms of the frequency response.

METHODS OF SYNTHESIS

Most of the synthesis methods currently in use employ cut and try methods to realize performance specifications. This is not very efficient and frequently the most feasible arrangement of parameters is not found because of the time absorbing details.

Often it is difficult to obtain an over-all perspective of system performance. Special computing aids and charts are usually used to speed synthesis,

*This paper was presented at the 3rd Southwest Conference on Feedback Control Systems in Dallas, Texas, April 28, 1955. The conference was sponsored by the Department of Electrical Engineering, Southern Methodist University in cooperation with the North Texas Section, A.I.E.E., the Professional Group on Automatic Control of the Dallas-Ft. Worth Section, IRE.

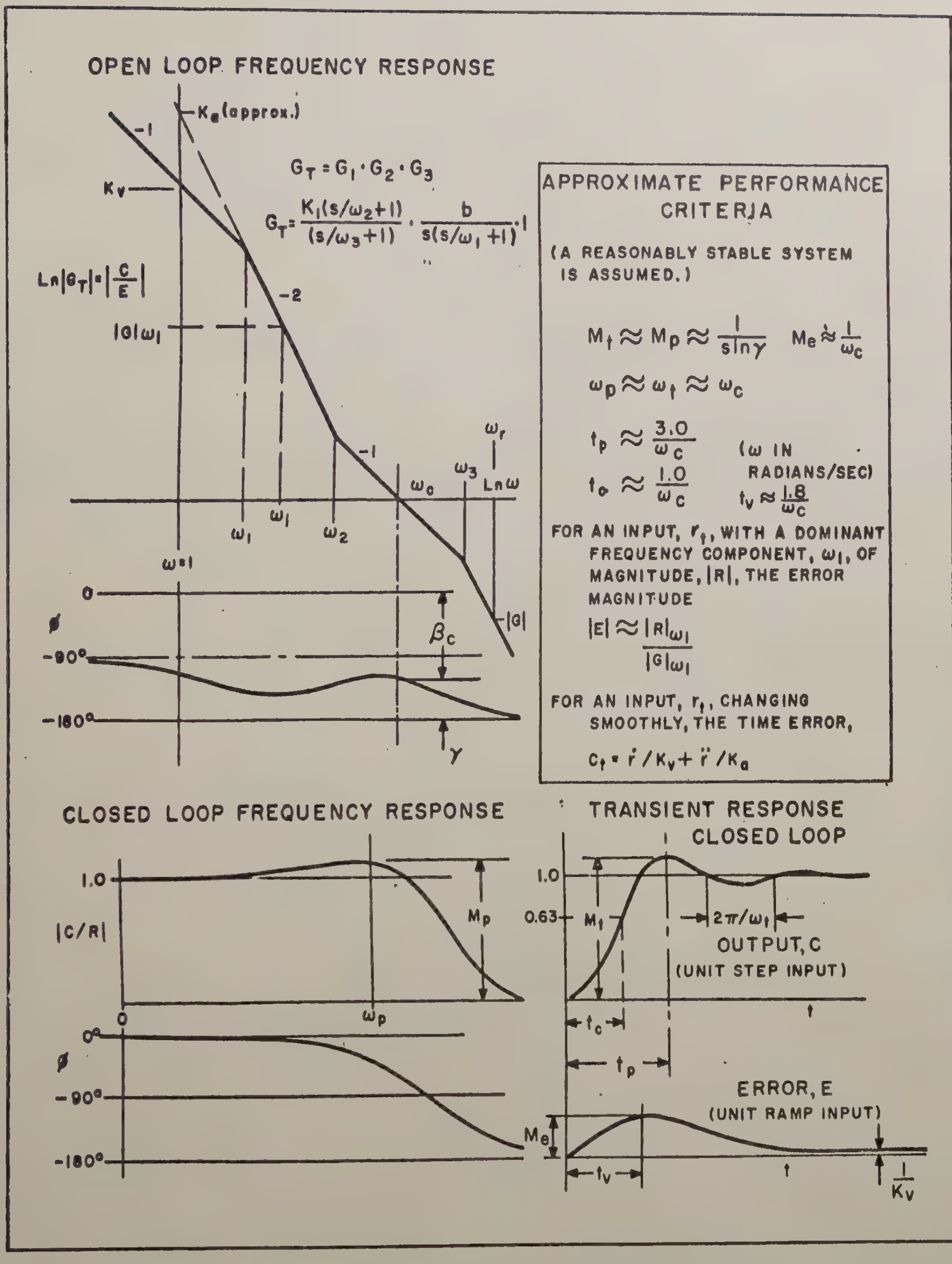


Fig. 1 - Approximate performance criteria from the frequency response.

but they are not always available, and they do not always show basic relationships between system parameters which could improve design. However, charts and computing aids have an important part in system design, particularly in analysis and in checking the quciker, more direct, but less accurate method to be described. Although it is sometimes inaccurate (there is a warning when that inaccuracy is present, however), it is a powerful addition to the many analysis and synthesis techniques now in use. It can be used to obtain gen-

eral relationships between system parameters which uniquely optimize the system to meet design specifications.

From simple, approximate equations, the necessary system compensation and transfer function may be directly determined, once the performance is specified, in terms of system gain, band-pass and degree of stability. This synthesis technique will be referred to as "asymptotic synthesis."

"MINIMUM LEAD" SYSTEMS

Because an infinite variety of feedback control systems could be considered, the discussion will be limited to a type of system, described below, common to many high performance servomechanisms.

It may be noted that high performance systems often have large integral networks or double integrators, a high gain at low frequencies, and a minimum phase angle, β_m, which occurs near the crossover frequency. This is illustrated in Fig. 2.

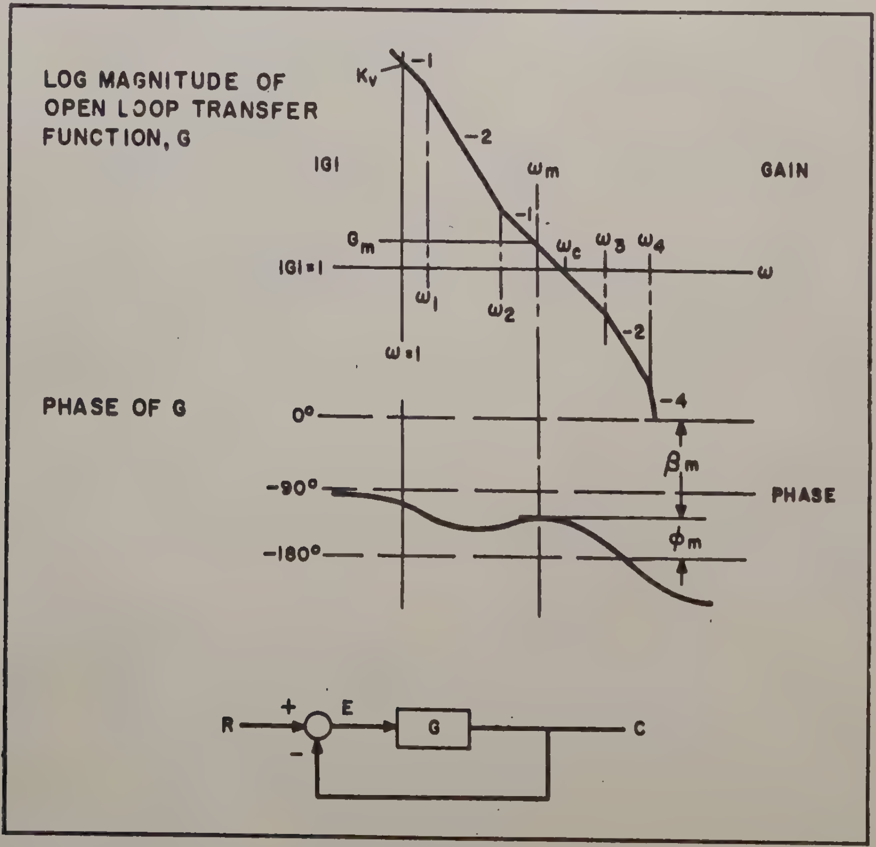


Fig. 2 - Asymptotic gain and phase angle of G.

The minimum phase angle occurs at a frequency ω_m, below the open loop crossover frequency, ω_c; G_m, the open loop gain at ω_m, equals ω_c/ω_m, assuming that slope of the open loop magnitude is -1 where the open loop magnitude is unity. For a given M_p, the maximum magnitude of the closed loop response, it would be desirable to have the frequency response locus of G, when plotted on

a Nichols chart, tangent to the given M circle at its maximum excursion from 180°. This is indicated in Fig. 3. The gain at this point is an optimum for a given M_p because a higher or lower gain will produce a larger M_p.

In the design, this is accomplished by placing ω_m at the desired tangent point. Thus, for a given M_p, ω_m should have the following relationship to ω_c:

$$\omega_m = \omega_c \cos \phi_m = \omega_c \frac{\sqrt{M_p^2 - 1}}{M_p} \quad ; \quad M_p = \frac{1}{\sin \phi_m}$$

These relationships are derived in the Appendix.

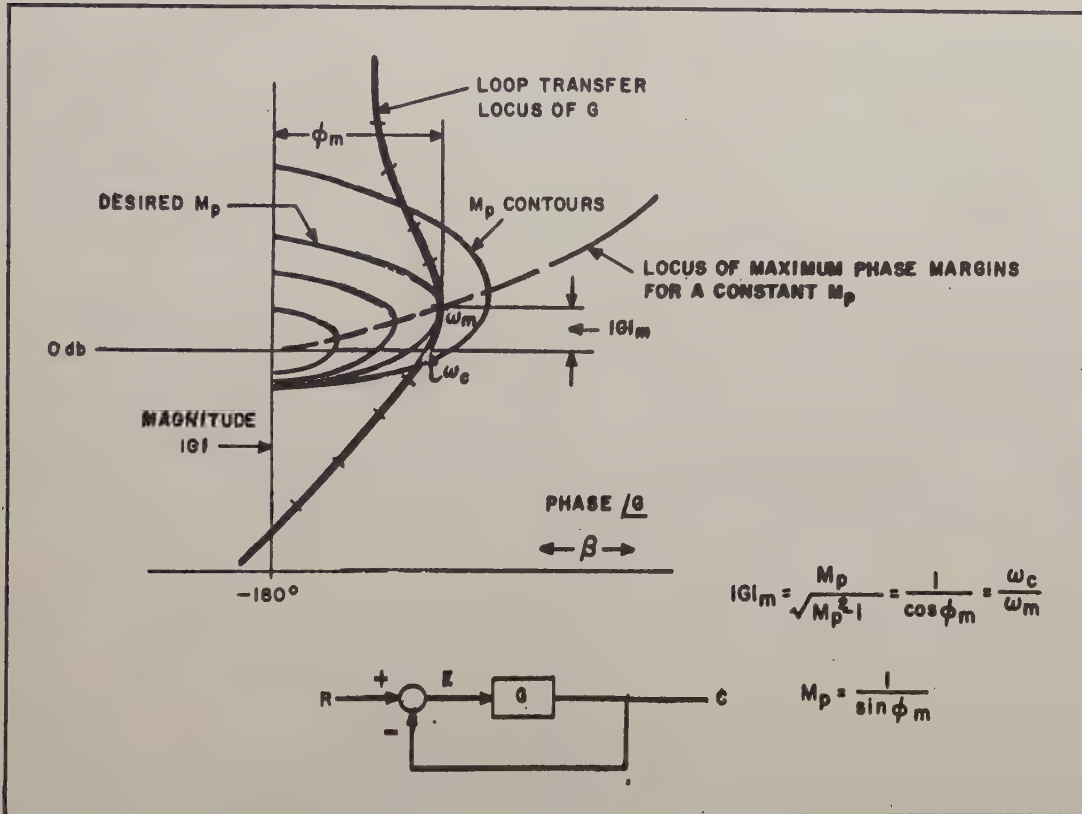


Fig. 3 - Nichols chart of minimum lead system.

Such an arrangement provides the desired stability with a minimum amount of effective lead (as denoted by the -1, or -6 db/octave, slope through ω_c) for a particular crossover frequency. This often economizes the amount of compensation needed for stability; it helps reduce noise within the system and optimizes the system gain. Furthermore, the transient settling time of such a system will be close to a minimum for a given band-pass.

A system with this design will be referred to as a "minimum lead" system.

ASYMPTOTIC SYNTHESIS METHODS

The method of asymptotic synthesis is relatively simple. The desired system transfer function, which appears to meet the most severe of the many

possible specifications, is sketched roughly to obtain the asymptotic form of the transfer function. Usually several of the corner frequencies will not be known, and it is desired that these corner frequencies be determined to provide the necessary, if not optimum, gain band-pass and degree of stability.

Three simple basic equations, and possibly others, are written. The asymptotic sketch of log gain vs. log frequency (log-log coordinates) is a valuable guide when writing the equations and when making the final check of the calculated system. These algebraic equations are written to relate:

1. the minimum phase angle of the open loop near the crossover frequency to the known and unknown corner frequencies of the system,

2. the system corner frequencies in a manner which will establish a "minimum lead" system,

3. the desired open loop gain to the known and unknown corner frequencies and to the system crossover frequency.

THE BASIC EQUATIONS

These sets of equations contain common unknowns, and the equations may be solved simultaneously. The method of deriving these equations will now be discussed, and applications of them will be given.

Equations relating the phase angle of the open loop near the crossover frequency (or between any corner frequencies) may be written as follows:

Consider the following open loop transfer function:

$$G = \frac{(1 + T_1 s)\ (1 + T_2 s)}{(1 + T_3 s)\ (1 + T_4 s)}$$

$$G\Big|_{j\omega_c} = \frac{\left(1 + \frac{j\omega_c}{\omega_1}\right)\left(1 + \frac{j\omega_c}{\omega_2}\right)\ \ldots\ldots}{\left(1 + \frac{j\omega_c}{\omega_3}\right)\left(1 + \frac{j\omega_c}{\omega_4}\right)^{\alpha}\ \ldots\ldots}$$

where: ω_c = frequency at which phase is desired, and
$\omega_1, \omega_2, \omega_3, \omega_4$ = corner frequencies.

From $G\Big|_{j\omega_c}$ the phase shift, β_c, may be written:

$$\beta_c = \text{arc tan}\left(\frac{\omega_c}{\omega_1}\right) + \text{arc tan}\left(\frac{\omega_c}{\omega_2}\right) + \ldots - \text{arc tan}\left(\frac{\omega_c}{\omega_3}\right)$$

$$-\alpha\,\text{arc tan}\left(\frac{\omega_c}{\omega_4}\right)\ \ldots\ldots$$

Each term on the right may be represented by an infinite series.

It would seem that the use of an infinite series in place of the arc tan of a number would infinitely complicate the problem of finding the phase shift, β_c, of $G\Big|_{j\omega_c}$. A very good approximation can be obtained, however, by using the first term of the series when $\omega_c/\omega_1 < 1$, and by using the first two terms of the series when $\omega_c/\omega_1 > 1$. Thus, arc tan $(\omega_c/\omega_1) \approx \omega_c/\omega_1$, when $\omega_c/\omega_1 < 1$ (corner frequency above ω_c); $\approx \pi/2 - \omega_1/\omega_c$, when $\omega_c/\omega_1 > 1$ (corner frequency below ω_c). (If a denominator quadratic should exist, the approximate phase shift at ω_c would be: $-2\zeta\,\omega_c/\omega_n$, $\omega_n > \omega_c$; or $-\pi + 2\zeta\,\omega_n/\omega_c$, $\omega_n < \omega_c$.) Note that the frequency ratio (of corner frequency and frequency at which phase is desired) is always less than 1 in the approximation.

This approximation is especially good if the frequency ratio in the approximation is equal to or less than 0.5. In other words, the approximation holds particularly well when the frequency at which the phase shift is desired is at least 2:1 from the nearest corner frequency. The crossover frequency is usually in such a position, and it is in this region that the phase shift is most important. Frequently the phase shift is specified in this region because it is here that the phase shift determines the degree of system stability.

The error due to the arc tan approximation is shown in Fig. 4. Correction factors could be used to minimize the error, but errors tend to cancel, and the method is usually accurate enough without them. Correction factors would only add complexity, and the advantages of this synthesis method would be decreased.

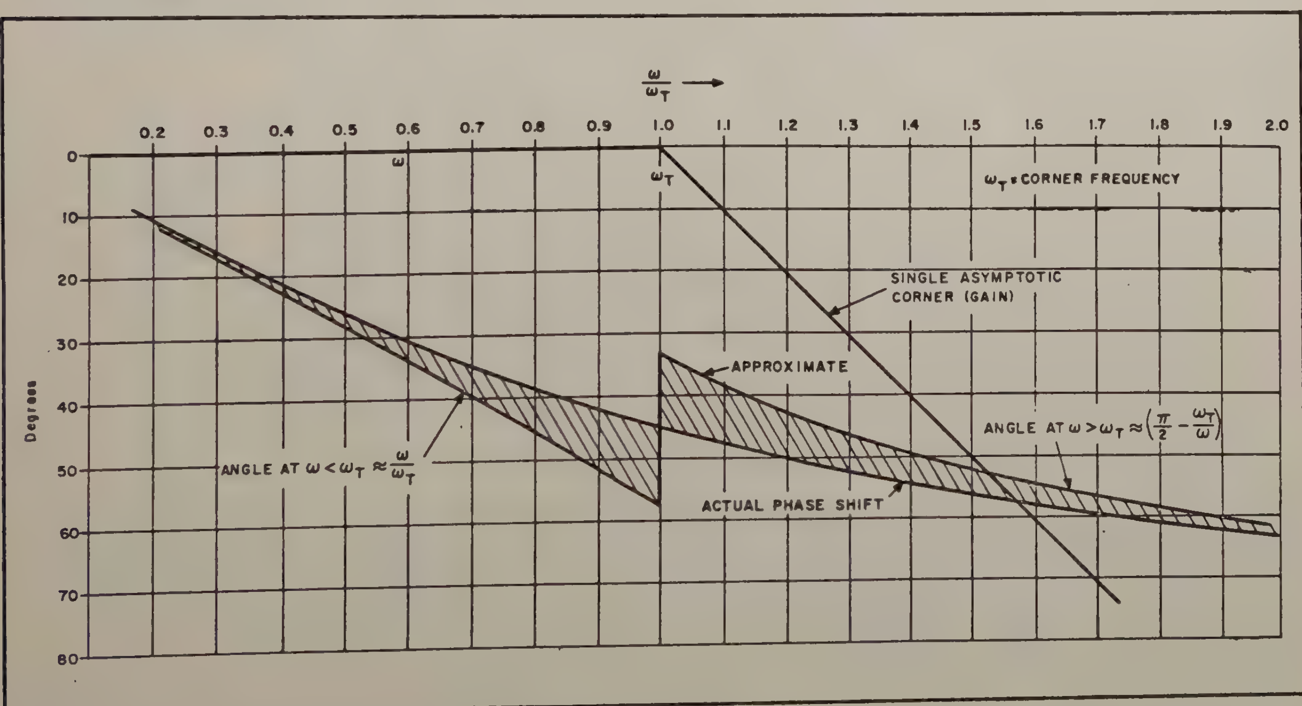


Fig. 4 - Comparison between actual and approximate phase shift for a single asymptotic corner as computed by conventional and approximate methods.

Thus, an equation for the phase shift can be reduced from a summation of trigonometric terms to algebraic terms. As shown in Fig. 5, an expression for the phase angle, β_c, at ω_c can be written from inspection of the asymptotic diagram, and

$$\beta_c = -\frac{\pi}{2} - \left(\frac{\pi}{2} - \frac{\omega_1}{\omega_c}\right) + \left(\frac{\pi}{2} - \frac{\omega_2}{\omega_c}\right) - \frac{\omega_c}{\omega_3} - 2\left(\frac{\omega_c}{\omega_4}\right).$$

Note that:

1. $-\pi/2$ occurs because of the -1 slope,

2. a term appears for each corner frequency,

3. each term is multiplied by plus or minus the change of slope at the corner frequency,

4. the corners may or may not be known.

Thus, β_c can be expressed as an algebraic equation

$$\beta_c = -\frac{\pi}{2} + \frac{\omega_1}{\omega_c} - \frac{\omega_2}{\omega_c} - \frac{\omega_c}{\omega_3} - 2\frac{\omega_c}{\omega_4}$$

This is the first of the three basic system equations.

The second basic equation is that which relates the system corner frequencies in a manner which will establish a "minimum lead" system. Of course, this may not always be required; other specifications may take precedence. This specification may unduly complicate the problem, or it may produce redundant equations if too many variables are specified.

This equation is found by differentiating the first equation for β_c. Thus,

$$\frac{d\beta}{d\omega_c} = -\frac{\omega_1}{\omega_c^2} + \frac{\omega_2}{\omega_c^2} - \frac{1}{\omega_3} - \frac{2}{\omega_4}$$

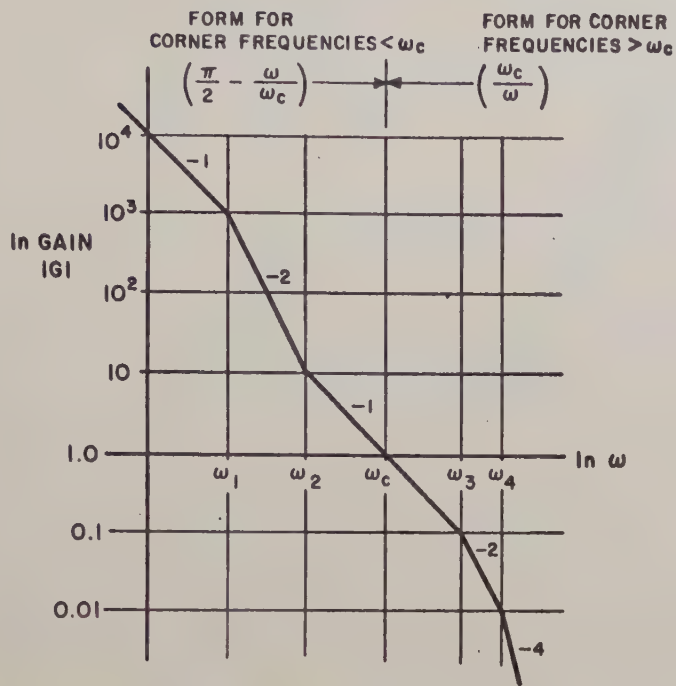
$$\frac{(\omega_2 - \omega_1)}{\omega_c^2} = \left(\frac{1}{\omega_3} + \frac{2}{\omega_4}\right)$$

Although ω_2, ω_1, or ω_c may be determined from the last equations in terms of other frequencies, it is usually more convenient, for later substitution, to separate the corner frequencies above and below ω_c as shown.

The third basic equation relates the desired open loop gains to corner frequencies, known and unknown.

This can be developed best by referring to Fig. 6.

Consider the gain between the frequencies ω_1 and ω_2. On log-log coordinates the gain change is a straight line with a slope of 2:1.



BY INSPECTION OF THE ASYMPTOTIC PLOT, THE PHASE ANGLE, β_c, AT ω_c IS APPROXIMATELY:

$$\beta_c = -\frac{\pi}{2} - \left(\frac{\pi}{2} - \frac{\omega_1}{\omega_c}\right) + \left(\frac{\pi}{2} - \frac{\omega_2}{\omega_c}\right) - \left(\frac{\omega_c}{\omega_3}\right) - (2)\left(\frac{\omega_c}{\omega_4}\right)$$

$2\omega_2 \leq \omega_c \leq \frac{\omega_3}{2}$ $\qquad$ β_c IS IN RADIANS

NOTE THAT CHANGE OF SLOPE AT CORNER FREQUENCY PRECEDES THE ARC TAN TERM.

PHASE MARGIN, γ, at $\omega_c = 180° + \frac{180}{\pi}\beta_c$

NOTE: FREQUENCY SCALE IS IN LN. RADIANS TO RELATE CORNER FREQUENCIES TO TIME CONSTANTS, AND THAT −1, −2, SLOPES ≅ −6, −12, db/OCTAVE

Fig. 5 - Approximate phase calculations.

Thus, by the definition of a -2 slope,

$$\frac{\log G_2 - \log G_1}{\log \omega_2 - \log \omega_1} = +2$$

or:

$$\frac{G_2}{G_1} = \left(\frac{\omega_2}{\omega_1}\right)^{+2}.$$

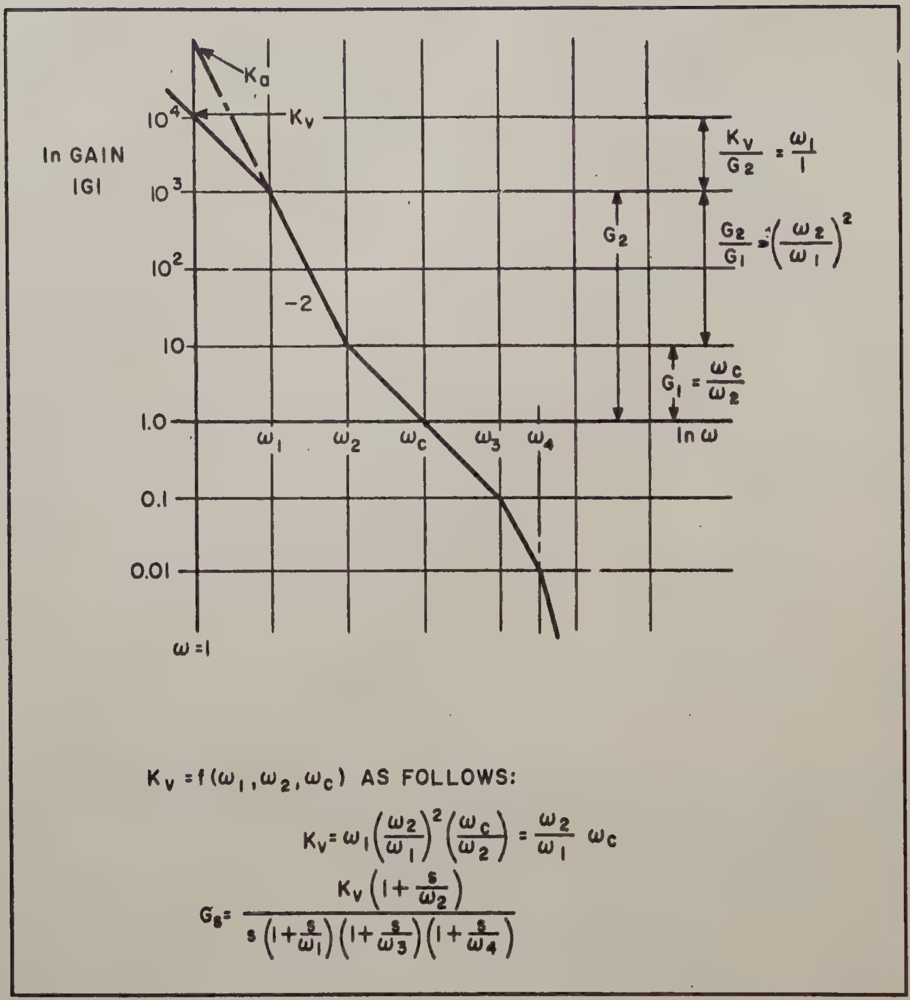


Fig. 6 - Relationships between gain and phase.

For $\mp$ n slope between ω_1 and ω_2,

$$\frac{G_2}{G_1} = \left(\frac{\omega_2}{\omega_1}\right)^{\pm n}.$$

Thus, by inspection of the asymptotic diagram in Fig. 6, the velocity constant of the system,

$$K_v = G_1 \left(\frac{G_2}{G_1}\right) \left(\frac{K_v}{G_2}\right).$$

$$K_v = \left(\frac{\omega_c}{\omega_2}\right) \left(\frac{\omega_2}{\omega_1}\right)^2 \left(\frac{\omega_1}{1}\right).$$

$$K_v = \left(\frac{\omega_2}{\omega_1}\right) \omega_c.$$

This algebraic equation must always hold true for the configuration shown in Fig. 6. It also serves as a check of the asymptotic drawing.

In a similar manner $K_a \approx (\frac{\omega_c}{\omega_2}) (\frac{\omega_2}{1})^2$

$$K_a \approx \omega_c \omega_2$$

and similarly, the asymptotic gain may be specified at any frequency as a function of the corner frequencies.

To illustrate the use of these equations, consider a typical system in the following problem:

Suppose that a sinusoidal input with a frequency of 1 radian/second and an amplitude of 14° peak is applied to a position system. It is desired that the peak sinusoidal error be less than 0.05° peak. It is also required that the lowest possible system gain and band-pass be used and that the frequency response peak magnitude be equal to or less than 1.3. It is known that a single integration exists, and that a corner frequency occurs at 180 rad/sec. It is desired to attenuate frequencies above the crossover as soon as possible.

The form of the open loop function that will meet the requirements is shown in Figs. 2 and 3 except that ω_4 will eqaul 180 rad/sec, and it will consist of a single corner frequency. ω_1 will be placed at $\omega = 1$ rad/sec, and the asymptotic gain at $\omega_1 = 1$ (at the corner frequency) will be $\sqrt{2}\,(14/.05) = 397 \approx 400$. This is also the velocity constant.

From the specifications, it will be necessary to have a "minimum lead" system as previously defined.

This means that the minimum phase shift in the system must occur below ω_c at ω_m, where:

$\omega_m = \omega_c \cos \phi_m$,

ω_m is the frequency at which the minimum phase β_m occurs, and

$\phi_m = 180^\circ + \beta_m$.

The basic calculations, the three approximate equations, are derived in Fig. 7. They are also shown below:

1. Phase shift at ω_m:

$$\beta_m = -\frac{\pi}{2} + \frac{(\omega_1 - \omega_2)}{\omega_m} - \omega_m \left(\frac{1}{\omega_3} + \frac{1}{180}\right)$$

2. Condition that makes β_m a minimum:

$$\frac{\omega_2 - \omega_1}{\omega_m^2} = \frac{1}{\omega_3} + \frac{1}{180}$$

2. Relation between K_v and corner frequencies:

$$K_v = \omega_c \frac{\omega_2}{\omega_1}$$

Three other basic equations, derived in the Appendix, complete the mathematical description of the system. For a minimum lead system, these relationships are exact.

$$\sin \phi_m = 1/M_p$$

$$\omega_m = \omega_c \cos \phi_m$$

$$\phi_m = \quad + \beta_m$$

THE FIRST BASIC EQUATION:

(a)

$$\beta_m = -\frac{\pi}{2} - \left(\frac{\pi}{2} - \frac{\omega_1}{\omega_m}\right) + \left(\frac{\pi}{2} - \frac{\omega_2}{\omega_m}\right) - \frac{\omega_m}{\omega_3} - \frac{\omega_m}{180}$$

$$\beta_m = -\frac{\pi}{2} - \frac{1}{\omega_m}(\omega_2 - \omega_1) - \omega_m\left(\frac{1}{\omega_3} + \frac{1}{180}\right)$$

(THE ANGLE AT ω_m, NOT ω_c, IS SPECIFIED)

THE SECOND EQUATION:

(b)

$$\frac{d\beta_m}{d\omega_m} = \frac{\omega_2 - \omega_1}{\omega_m^2} - \left(\frac{1}{\omega_3} + \frac{1}{180}\right)$$

$$\frac{\omega_2 - \omega_1}{\omega_m^2} = \frac{1}{\omega_3} + \frac{1}{180}$$

(NOTE THAT THE FREQUENCIES ABOVE AND BELOW ω_c ARE SEPARATED)

THE THIRD EQUATION:

(c)

$$K_v = \frac{\omega_c}{\omega_2}\left(\frac{\omega_2}{\omega_1}\right)^2 \omega_1$$

$$K_v = \frac{\omega_c \omega_2}{\omega_1}$$

$$\sin \phi_m = \frac{1}{M_p} \qquad \omega_m = \omega_c \cos \phi_m \qquad \phi_m = \pi + \beta_m$$

THESE EQUATIONS ARE GENERAL, AND WITH STUDY, USEFUL INFORMATION CAN BE OBTAINED FROM THEM. THERE ARE 9 UNKNOWNS; 6 EQUATIONS. IF 3 PARAMETERS ARE SPECIFIED, THE REMAINING 6 WILL BE DETERMINED.

Fig. 7 - Asymptotic synthesis equations.

These equations are quite general and with study useful design formulas can be obtained from them. There are nine unknowns, six equations for the system shown in Fig. 2. If any three parameters are specified, the remaining six can be determined by solving the equations simultaneously. This may be done rapidly and the results are surprisingly accurate if

$$2\omega_2 \leq \omega_m \leq \omega_3/2.$$

For the problem in question, three variables have been determined from the specifications:

$$\omega_1 = 1 \qquad\qquad M_p = 1.3 \qquad\qquad K_v = 400$$

These values have been substituted into the three equations and the simultaneous solution is given in Fig. 8. From the solution,

$$\begin{aligned}\omega_2 &= 9.9 \text{ rad/sec}\\ \omega_c &= 40.5 \text{ rad/sec}\\ \omega_m &= 25.9 \text{ rad/sec}\\ \omega_3 &= 129.5 \text{ rad/sec}\end{aligned}$$

and the open loop transfer function, G(s) becomes:

$$G = \frac{400\left(\frac{s}{10} + 1\right)}{s(s+1)\left(\frac{s}{130} + 1\right)\left(\frac{s}{180} + 1\right)}$$

DISCUSSION

1. Accuracy of Calculations

The equations in Fig. 8 constitute the complete calculations. The six unknowns were uniquely determined. Note that $2\omega_m = 51.8$, $\omega_m = 12.95$, $\omega_3 = 129.5$ and $\omega_2 = 9.9$. This shows that the condition $2\omega_2 \leq \omega_m \leq \omega_3/2$ has been met, and that the approximate phase equation (1) should have been quite satisfactory.

Only a small error in these first calculations should be expected. Certainly the calculations are more accurate than the system parameters can be measured. Nevertheless, no approximate method should be trusted. Perhaps no one solution should be judged as final until it has been checked by another method. This solution was checked for possible revision by graphical methods.

The corner frequency-gain relationships were checked with an asymptotic plot. This is shown in Fig. 9.

The M_p and ω_m were checked by substituting $s = j\omega$ in the transfer function and calculating the actual $G\Big|_{j\omega}$. This was plotted on a Nichols chart. The results are reproduced in Fig. 10. This was the first calculation.

FOR THE PROBLEM IN QUESTION, 3 PARAMETERS ARE SPECIFIED:

$\omega_1 = 1 \qquad M_p = 1.3 \qquad K_v = 400$

THEN FROM FIGURE 7,

$\phi_m = 50.3° = 0.88$ RADIANS; $\beta_m = -129.7° = -2.26$ RADIANS; $\omega_m = 0.64\,\omega_c$.

USE EQ. (c) $\left[K_v = \frac{\omega_c \omega_2}{\omega_1}\right]$;

$$\omega_c = \frac{400}{\omega_2} \qquad (1)$$

USE EQ. (a) $\left[\beta_m = -\frac{\pi}{2} - \frac{\omega_2 - \omega_1}{\omega_m} - \omega_m\left(\frac{1}{\omega_3} + \frac{1}{180}\right)\right]$

$$-2.26 = \frac{\pi}{2} - \frac{(\omega_2 - 1)}{0.64\omega_c} - 0.64\,\omega_c\left(\frac{1}{\omega_3} + \frac{1}{180}\right)$$

$$0.688 = \frac{\omega_2 - 1}{0.64\omega_c} + 0.64\,\omega_c\left(\frac{1}{\omega_3} + \frac{1}{180}\right) \qquad (2)$$

COMBINE EQUATIONS (1), (2), (b); Eq. b, $\left[\frac{\omega_2 - \omega_1}{\omega_m} = \frac{1}{\omega_3} + \frac{1}{180}\right]$

$$0.688 = \frac{(\omega_2 - 1)(\omega_2)}{(0.64)(400)} + 0.64\,\frac{(400)}{\omega_2}\left(\frac{\omega_2 - \omega_1}{\omega_m^2}\right) = \frac{\omega_2(\omega_2 - 1)}{(0.64)(400)} + \frac{(\omega_2 - \omega_1)(\omega_2)}{(400)(0.64)}$$

$$\frac{(0.688)(0.64)(400)}{2} = \omega_2^2 - \omega_2 = 88$$

$$\boxed{\omega_2 = \frac{1 \pm \sqrt{1 + 4(88)}}{2} = 9.9 \text{ RAD/SEC}}$$

$\boxed{\omega_c = 40.5 \text{ RAD/SEC}}$ EQ. (1)

$\boxed{\omega_m = 25.9 \text{ RAD/SEC}}$ $\omega_m = 0.64\,\omega_c$

FROM Eq.(b), $\frac{(8.9)}{(25.9)^2} - \frac{1}{180} = \frac{1}{\omega_3}$ $\boxed{\omega_3 = 129.5 \text{ RAD/SEC}}$

AND THE SYSTEM OPEN LOOP FUNCTION $\Big\}$ $G = \frac{400}{s} \; \frac{\left(\frac{s}{10} + 1\right)}{(s+1)\left(\frac{s}{130} + 1\right)\left(\frac{s}{180} + 1\right)}$

Fig. 8 - Solution of equations in Fig. 7.

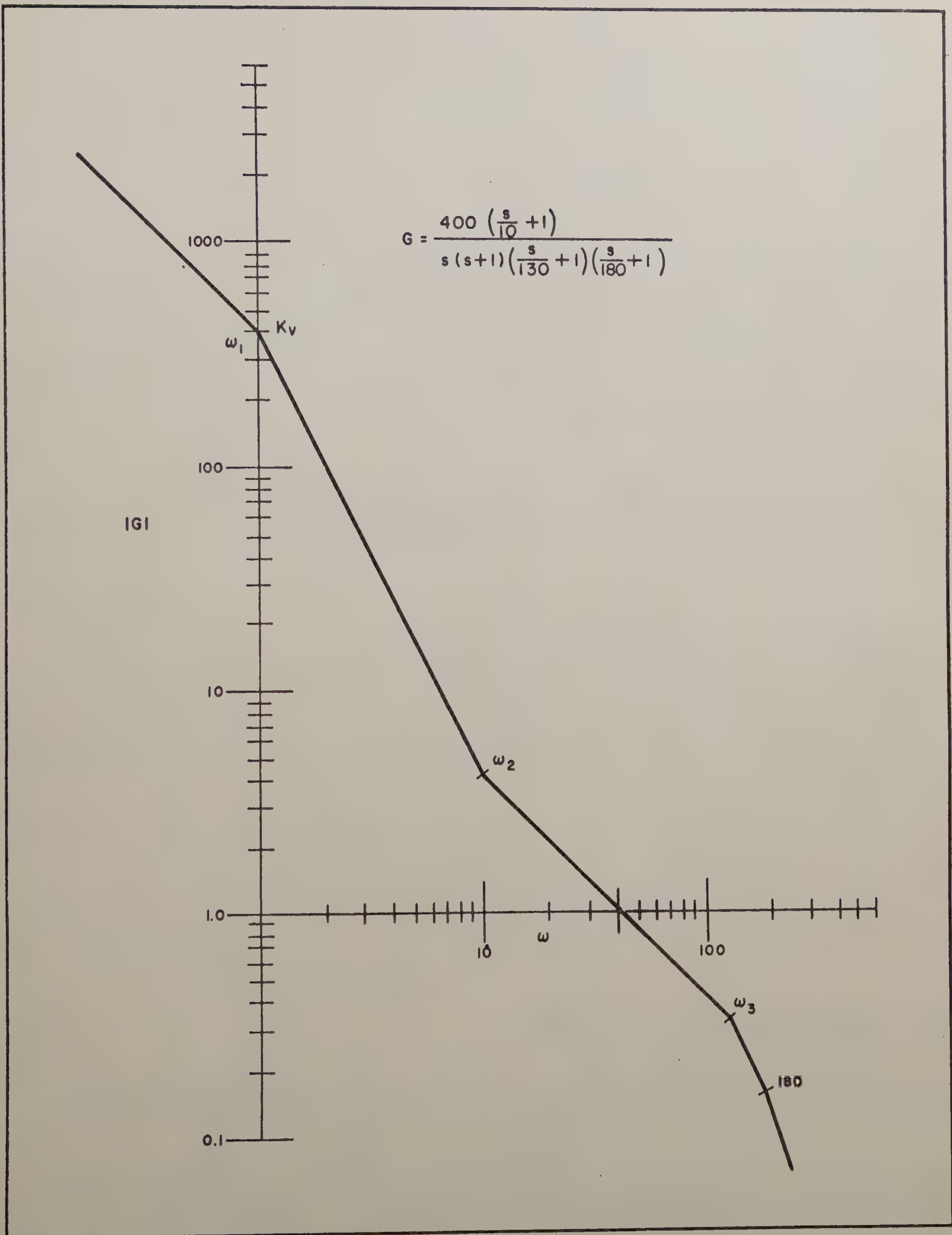


Fig. 9 - Asymptotic plot of synthesized system.

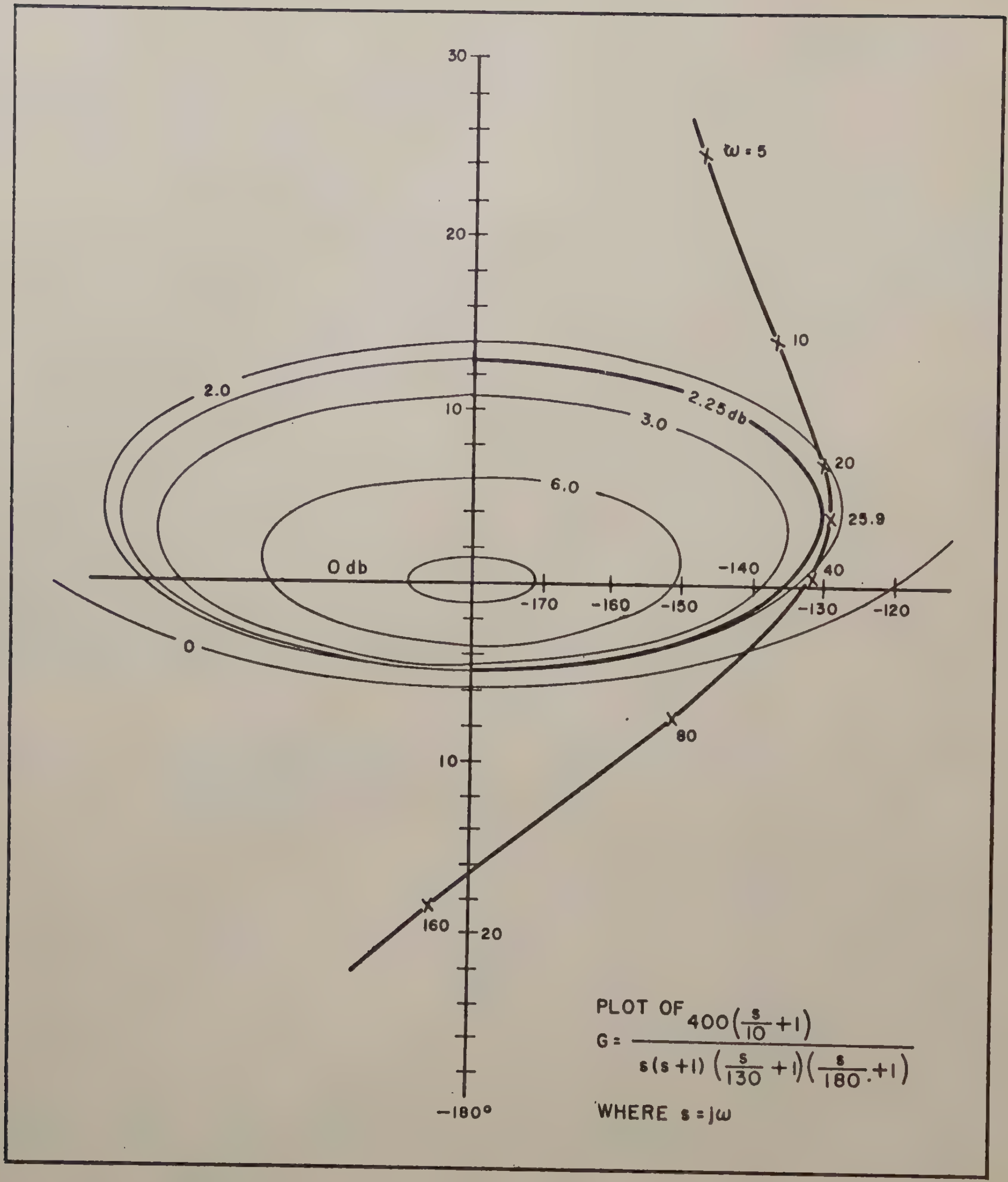


Fig. 10 - Nichols chart of synthesized system.

2. Limitations in Using the Method

Unfortunately, all of the uses, advantages and disadvantages of this synthesis method cannot be included in this brief discussion. Examples have served to illustrate its advantages and its scope.

Disadvantages of this method exist largely when it is improperly used. For instance, a complex system may yield an excessive number of equations, and the simultaneous solution may result in an equation of high order. Considerable time could be spent solving the equation for its roots, only one of which would be useful. Possibly a trial and error solution could be made more

quickly, but it would have to satisfy very closely the original simultaneous equations, written from the asymptotic diagram. Actually, the equations can usually be solved rather quickly because the allowable range of each corner frequency can be observed from the asymptotic diagrams, and considerable simplification can be made if some of the variables are known to be much larger than others, or if the range of variation is known. Thus, even for complex systems, asymptotic synthesis can be used advantageously.

Another apparent disadvantage in the method is that solutions may yield negative frequencies, complex frequencies which were not expected or corner frequencies which change the shape of the original asymptotic diagram. These solutions would be incorrect, but this is not usually a disadvantage or a detriment in the method; they merely indicate that the proposed system cannot meet the design specification, or that more restrictions have been placed on the system than the number of variables allow. It is necessary in these cases to reshape the asymptotic diagrams, reduce the severity of design specifications or compromise the requirements to prevent conflicting or redundant equations.

These modifications can be made quickly using asymptotic synthesis procedures, an over-all perspective of the possible system performance can be retained and the system design can be readily established.

3. Usefulness of the Method

When properly used and when its limitations are realized, the method of asymptotic synthesis can be a powerful tool to develop rapidly over-all relationships between system parameters and to establish a framework for optimum system design. It does not replace more detailed or more accurate studies; rather it supplements them. It reduces the problem of loop design to rather simple algebra which yields fairly accurate solutions directly with a minimum of trial and error.

APPENDIX

Derivation of the Locus of Maximum Phase Margin for a Given M_p on a Nichols Chart

$$M_p = \left| \frac{G}{1 + G} \right|$$

$$G = |G| \, \epsilon^{+j\beta}$$

$$\beta = -\pi + \phi_m$$

(Refer to Fig. 3 for symbols)

$$M_p = \left| \frac{1}{1 + \frac{1}{G}} \right| = \left| \frac{1}{1 + \frac{\epsilon^{-j\beta}}{|G|}} \right|$$

$$M_p = \left| \frac{1}{1 + \frac{1}{|G|} \varepsilon^{-j(-\pi + \phi_m)}} \right|$$

$$M_p = \left| \frac{1}{1 + \frac{1}{|G|} \left[\cos(\pi - \phi_m) + j \sin (\pi - \phi_m) \right.} \right|$$

$$M_p = \frac{1}{1 + \frac{1}{|G|} (-\cos \phi_m + j \sin \phi_m)}$$

$$M_p^2 = \frac{|G|^2}{|G|^2 - 2 |G| \cos \phi_m + 1}$$

For a given M_p:

$$|G|^2 - 2 |G| \cos \phi_m + 1 = \frac{G^2}{M_p^2}$$

$$\phi_m = \cos^{-1} \frac{|G|^2 M_p^2 + M_p^2 - |G|^2}{2 |G| M_p^2} = \cos\text{-}1 (X)$$

$$\frac{d\phi_m}{d G} = \frac{-1}{1 - X^2} \frac{dX}{dG}$$

For maximum ϕ_m, $\frac{dX}{dG} = 0$ and

$$|G| (2 |G| M_p^2 - 2 |G|) = (|G|^2 M_p^2 + M_p^2 - |G|^2)$$

$$2 |G|^2 M_p^2 - 2 |G|^2 = |G|^2 M_p^2 + M_p^2 - |G|^2$$

$$|G|^2 M_p^2 - M_p^2 = |G|^2$$

$$|G| = \frac{M_p}{\sqrt{M_p^2 - 1}} \qquad (2)$$

From Eq. (1), the maximum ϕ_m becomes:

$$\cos \phi_m = \frac{\dfrac{M_p^2}{M_p^2 - 1} + 1 - \dfrac{1}{(M_p^2 - 1)}}{\dfrac{M_p}{2\sqrt{M_p^2 - 1}}}$$

$$\cos \phi_m = \frac{(M_p^2 + M_p^2 - 1 - 1)\sqrt{M_p^2 - 1}}{2\, M_p\, (M_p^2 - 1)}$$

$$\cos \phi_m = \frac{\sqrt{M_p^2 - 1}}{M_p}\ ; \quad \sin \phi_m = \sqrt{1 - \cos^2 \phi_m} \tag{3}$$

$$\sin \phi_m = \frac{1}{M_p} \tag{4}$$

PERFORMANCE OF DRIVE MEMBERS IN FEEDBACK CONTROL SYSTEMS*

F. M. Bailey
General Electric Company
Schenectady, New York

Summary -- Limitations of system performance due to a mechanical structure are discussed, and practical methods of improving the performance of fire control systems with this limitation are illustrated.

The history of high performance feedback control systems has been closely associated with the development of naval gun power drives. Progress in this field had its start with the development of early thyratron, hydraulic and then amplidyne systems, and history has shown a preference for each depending upon the degree to which the components have been perfected at any given time.

Closely allied with this work has been the progressive development of fire control systems, an important element of which is the director. The feedback control problems of the director are in many ways similar to that of the gun mount drive, but it is the significant differences which are to be discussed in this paper.

PERFORMANCE REQUIREMENTS

In discussing performance we must first examine the requirements which a good system must meet. In reviewing this, the discussion will be restricted to the use of opened and closed loop frequency response diagrams which form the basis of system response criteria.

A simplified closed loop feedback system is shown in Fig. 1. It contains an amplifier, a power boost element and a driving element such as an electric motor or hydraulic piston. The driving unit actuates the load and is, in addition, connected mechanically to a selsyn control transformer which generates a difference signal between load position and order position. This signal energizes the servo amplifier.

Frequency response measurements may be made on the system by cutting the loop at the point (X) and measuring output shaft motion with an oscilloscope while introducing a constant amplitude variable frequency signal at the input. The open loop diagram thus obtained is shown in Fig. 2. For constant input, the output will drop off at the rate of 6 db per octave until the first time constant is reached (usually at the motor constant), at which point the curve will begin to attenuate at 12 db per octave and so on at an increased rate of 6 db per octave for each successive time constant. If the loop is closed with a gain represented by the magnitude A, the zero db axis will cross the curve

*This paper was presented at the 3rd Southwest Conference on Feedback Control Systems in Dallas, Texas, April 28, 1955. The conference was sponsored by the Department of Electrical Engineering, Southern Methodist University in cooperation with the North Texas Section, A.I.E.E., the Professional Group on Automatic Control of the Dallas-Ft. Worth Section, IRE.

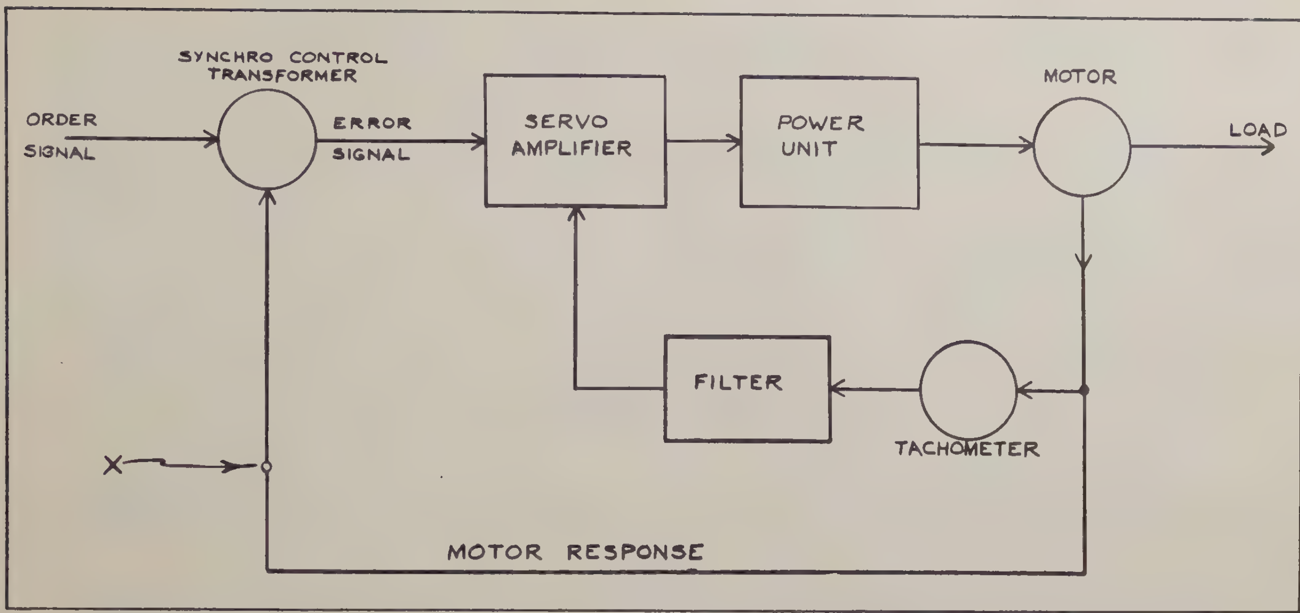
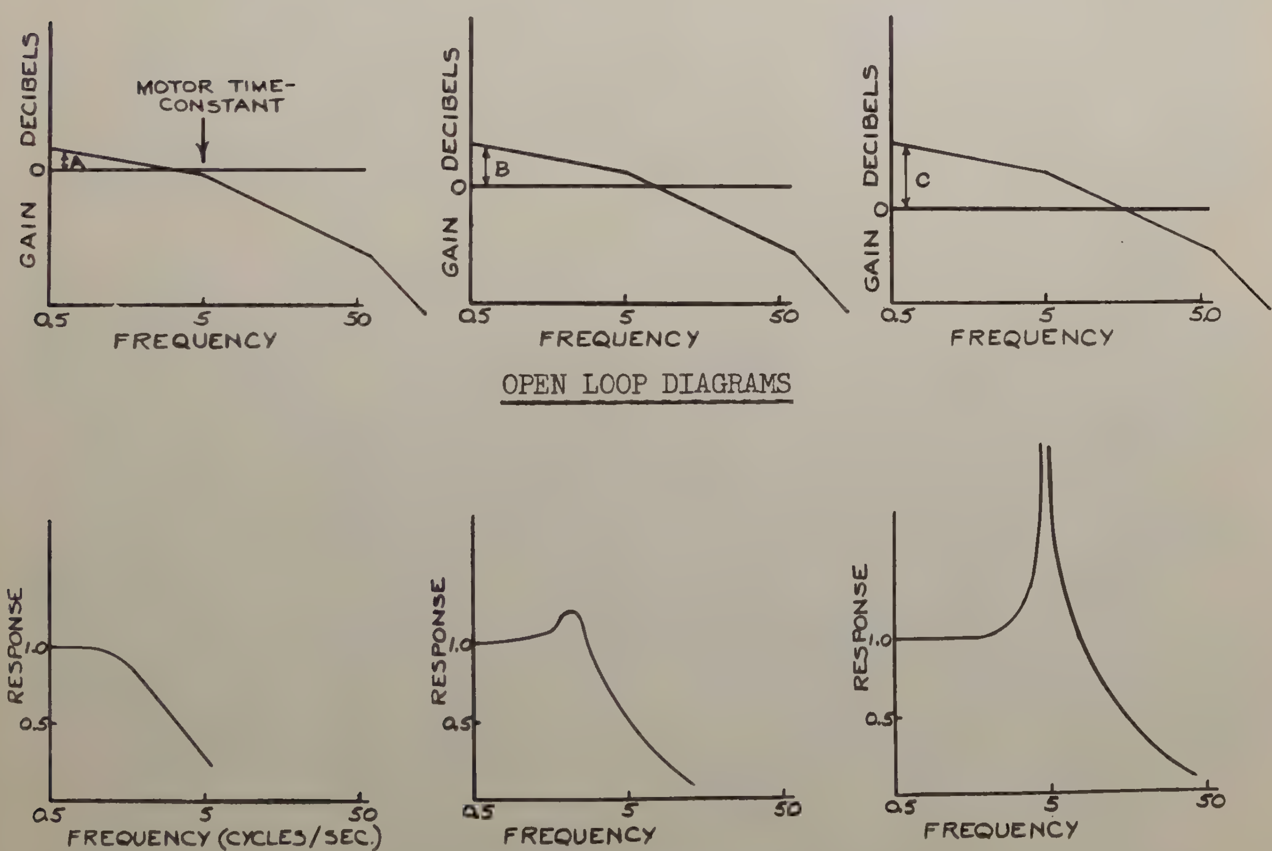


Fig. 1 - Servo control loop.



CLOSED LOOP DIAGRAMS

Fig. 2

at 6 db per octave and the system will be stable. The corresponding closed loop frequency response is shown in Fig. 2.

If the gain is increased as represented by magnitude B, the zero db axis will cross in a region close to 12 db per octave and the corresponding closed loop curve in Fig. 2 will show considerable resonance. If this process is extended by increasing the gain to a magnitude C, it is clear that the system will become unstable and break into oscillation. This then is basically the servo problem -- to raise the gain sufficiently to extend the frequency response of the closed loop diagram and at the same time prevent the approach of instability in the enclosed feedback loop.

TYPICAL PERFORMANCE

With this brief background, Fig. 3 can be examined to obtain a feeling for the type of performance which can be obtained with currently available servo system components.

This figure is a composite of the frequency response and phase shift curves of three of our major power drive systems. Curve A is an 8 hp unit capable of delivering 40 hp peaks and is typical of equipment currently being used in the field. It will be noted that the frequency peak rises to nearly 1.6 and that at $2\frac{1}{2}$ cycles, there is approximately 50° of phase shift. Curve C is the response for a more recent production unit having a 25 hp rating with capabilities of 400 hp peak. This is a cascaded generator system and it will be noted that the phase shift at $2\frac{1}{2}$ cycles is also 50°. On a test stand, the power drive by itself is capable of performing with a phase shift of 30°. Curve B shows the performance of an ignitron drive in the same power range as system number one. In this case, the phase shift at $2\frac{1}{2}$ cycles has been brought down to 25°.

The dotted line connecting the two humps in the response curve represents the performance achieved with the unit on a test stand in which a flywheel is used to simulate the load (instead of gearing and mount inertia). It will be noted that the mount resonance shows up in reality as an antiresonance. This is because the position loop is closed around the drive motor by means of additional instrument gearing and the gun mount is coupled through the gearing to the servo system. Experience has shown that closing the loop around a set of power gearing will lead to stability problems which seriously hamper the performance of the over-all equipment. The performance shown by these curves can be achieved only by the device of utilizing additional instrument gearing. It is also clear that in searching for increased servo performance, a limitation, even with this arrangement, is imposed by the mechanical structure utilized in the moving assembly. With equipment of reasonable mechanical configuration, and especially that hampered by the necessary emphasis on minimum weight, it seems unlikely that any appreciable system improvement can be achieved through servomechanism component improvement alone.

THE DIRECTOR PROBLEM

It has been pointed out that in gun drives it is virtually impossible to stabilize a high performance unit by taking the response from the one speed shaft, and that it is necessary to utilize instrument gearing from the motor

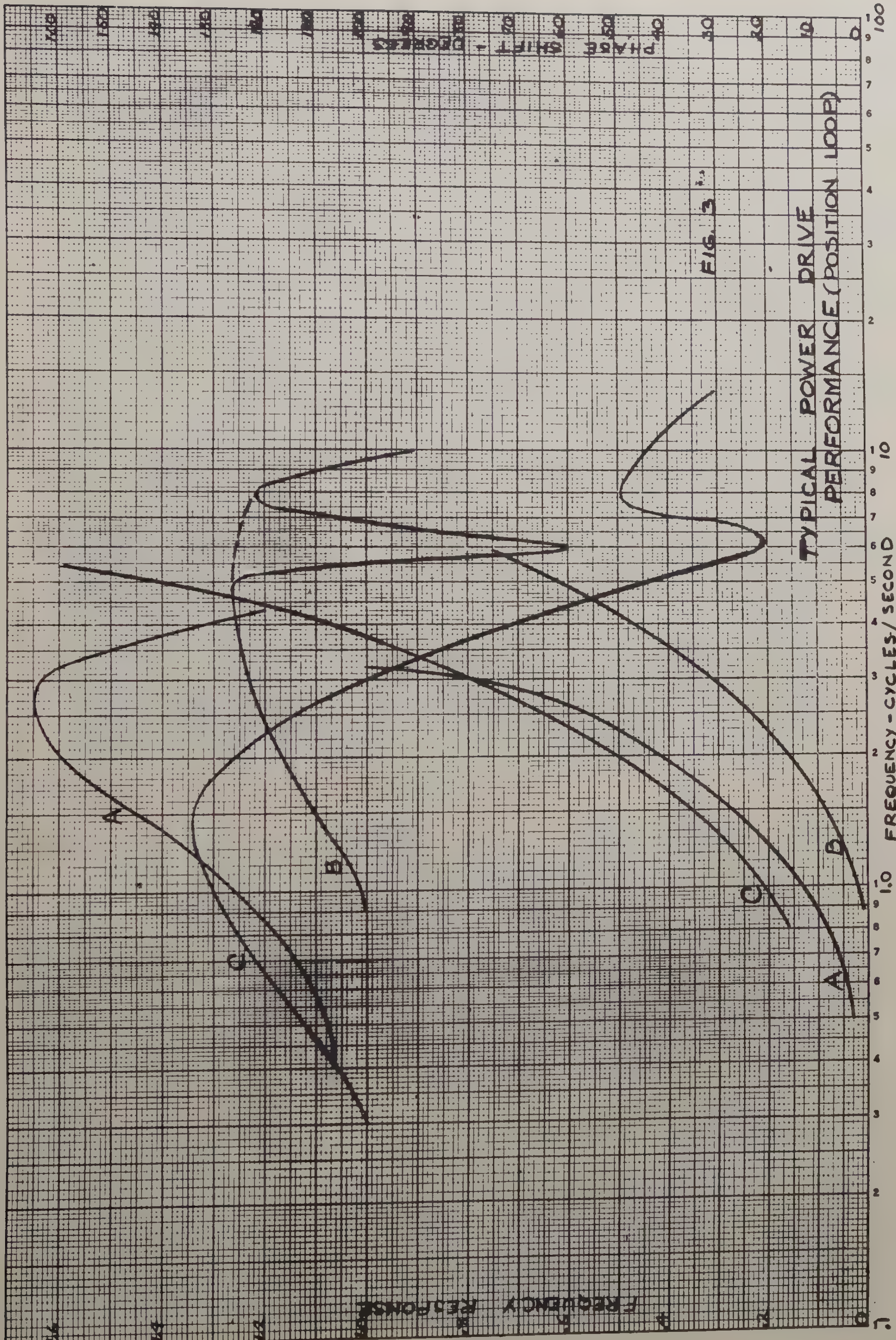

TYPICAL POWER DRIVE
PERFORMANCE (POSITION LOOP)
FIG. 3
FREQUENCY - CYCLES/ SECOND
PHASE SHIFT - DEGREES
FREQUENCY RESPONSE
A
B
C
D

Fig. 3

shaft for this purpose. A director drive, on the other hand, whose function is to force the radar antenna to track a target, must of necessity obtain its information from one speed motion since it is at this point that the antenna is connected. In this case, the power gearing is included in the closed loop and the additional time constants contribute to the stability problem. Resonant peaks for conventional units of this type come dangerously close to two and the phase shift involved may be well over 100° at resonance. Such performance is tolerable for a chain type of director system; but with continuing interest in the super-position of director and firing equipment, the resonant problems of the disturbed sight type of configuration become very critical. Here, in addition to the complexity of the gearing problem is added the additional complexity of an antenna servo all cascaded within the same control loop. The elimination of one or the other seems to be necessary in order to accomplish the end purpose.

Presently there are two methods of doing this. The first is to build an unwound director system in which a mechanically driven platform is mounted above the gun turret and is positioned by unwinding gearing. The second is to provide a power drive system in which there is no gearing. A discussion of the second system is within the scope of this paper.

Figure 4 shows a schematic diagram of the arrangement utilizing a one speed motor for driving either a director or gun platform as compared with a

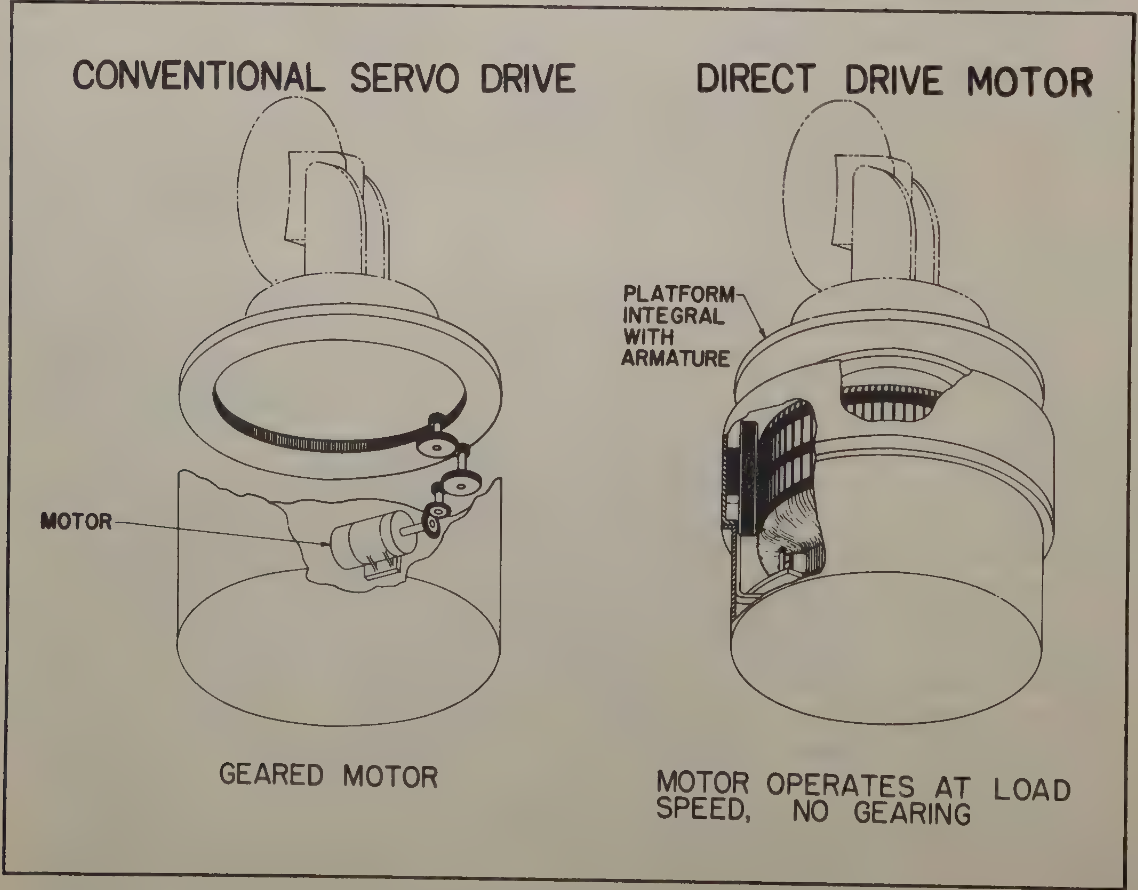


Fig. 4

Fig. 5

geared train system for doing the same thing. Figure 5 shows a photograph of an electric motor built for this purpose weighing 700 lbs. and providing an output torque of 1000 ft. lbs.

The frequency response of this unit is shown in Fig. 6. As far as can be ascertained, utilization of such a piece of equipment can make a disturbed line of sight director system successful. Such a unit employing a conventional style motor has been set up at the Schenectady County Airport and the tracking and computing operation accomplished utilizing a gyro lead angle computer.

It is realized that the weight considerations are entirely unsatisfactory for airborne equipment, and for this reason we have initiated a program to accomplish the same functional operation with hydraulic equipment.

Figure 7 presents a picture of a five piston hydraulic actuator which can be used as a one speed element in a power drive system. The building of such an equipment has been very recent and performance data is not available although the operation of similar hydraulic equipment leads us to be very optimistic.

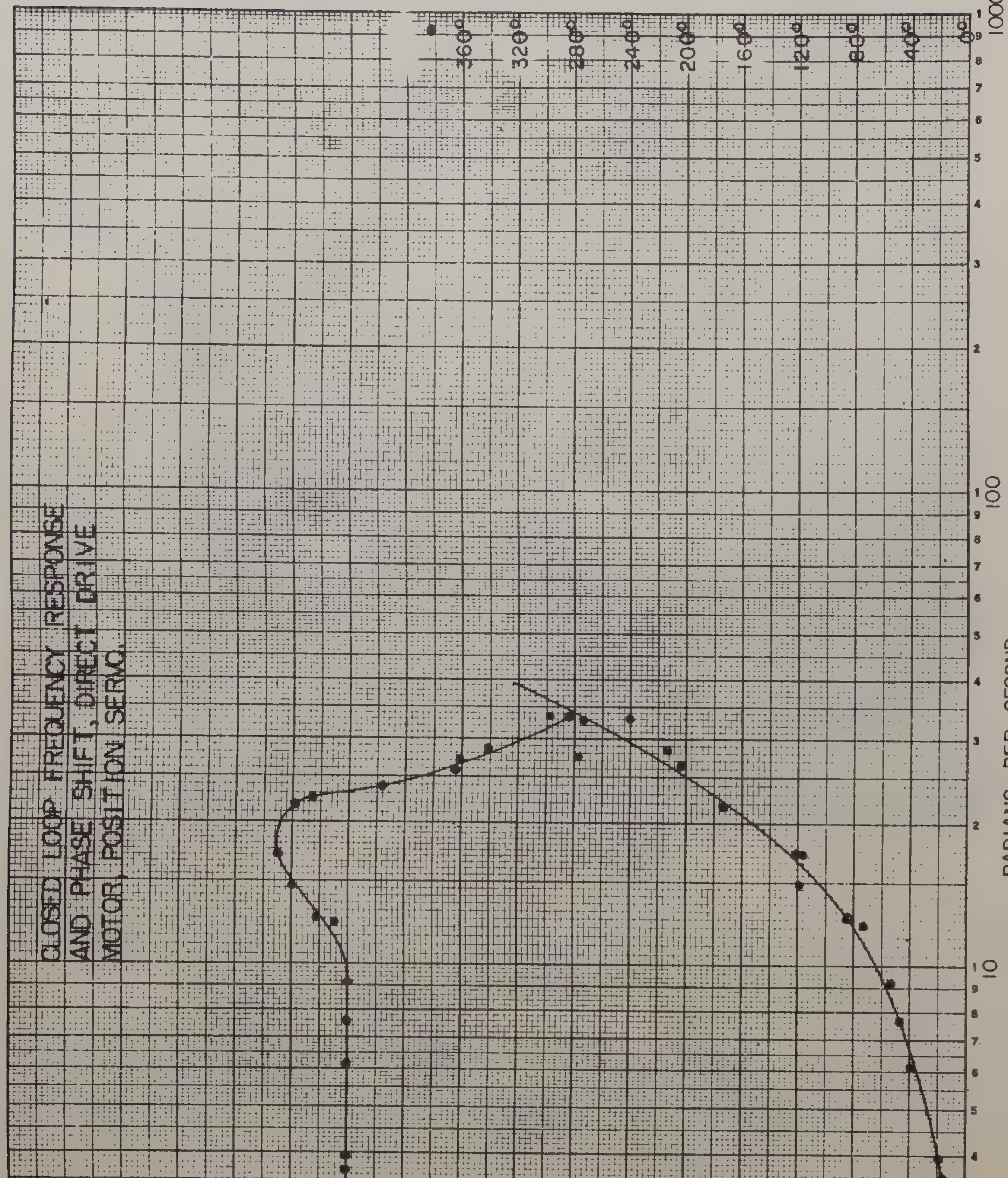

CLOSED LOOP FREQUENCY RESPONSE
AND PHASE SHIFT, DIRECT DRIVE
MOTOR, POSITION SERVO.
1.6
1.4
1.2
1.0
.8
.6
.4
.2
0
360°
320°
280°
240°
200°
160°
120°
80°
40°
0°
1
10
100
1000
RADIANS PER SECOND

Fig. 6

The previous discussion has demonstrated the benefits obtained from introducing simplicity in the feedback systems. The contributions in this area can be best provided by making components which, when combined, can form the simplest possible servo configuration. Since performance can be increased

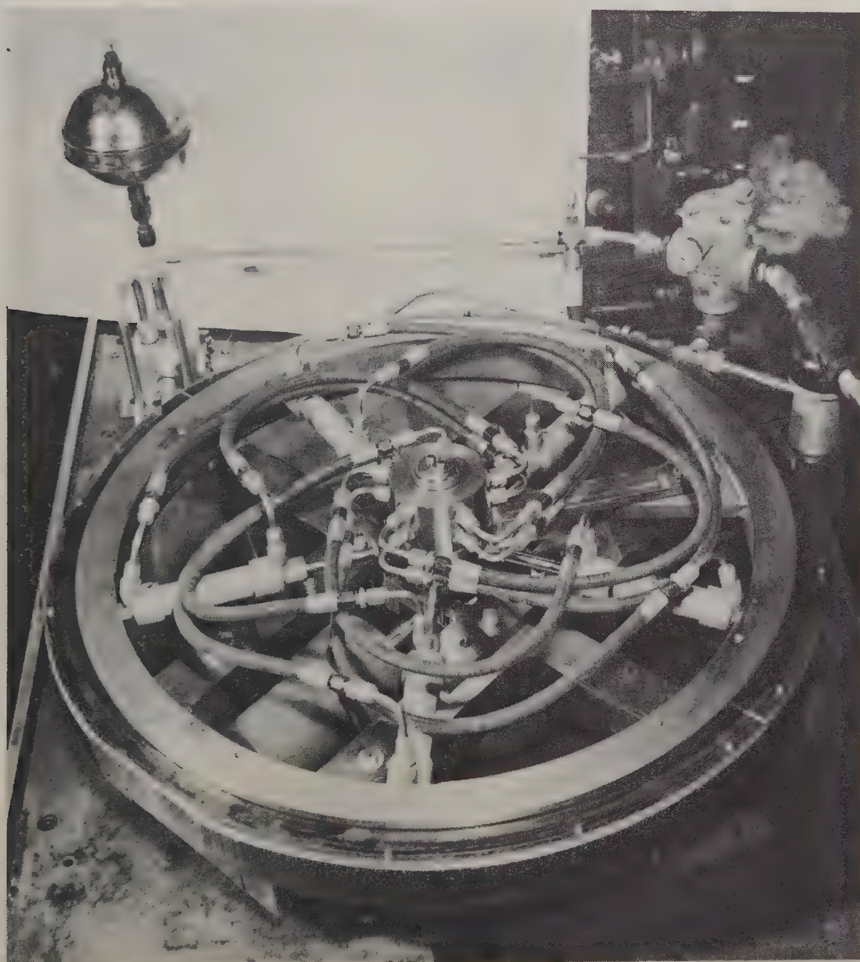


Fig. 7 - Five piston hydraulic rotary actuator.

by including a rigidly mounted tachometer, it is essential that we build actuating equipment such as servo motors with integrally mounted tachometers. For practical purposes nearly all servos require a brake, and the construction of a combination motor, tachometer and brake package is of advantage.

In Fig. 8 are shown two electric packages -- one a 2 hp, and the other a $\frac{1}{2}$ hp unit capable of delivering 25 ft. lbs. and 10 ft. lbs. of torque, respectively. A second step is to provide a power amplification unit which has a minimum number of time constants to be used to actuate the motor. In these horsepower ranges, such a system seems to be a generator since it offers the servo designer one time constant instead of two found in a normal amplidyne.

Fig. 8

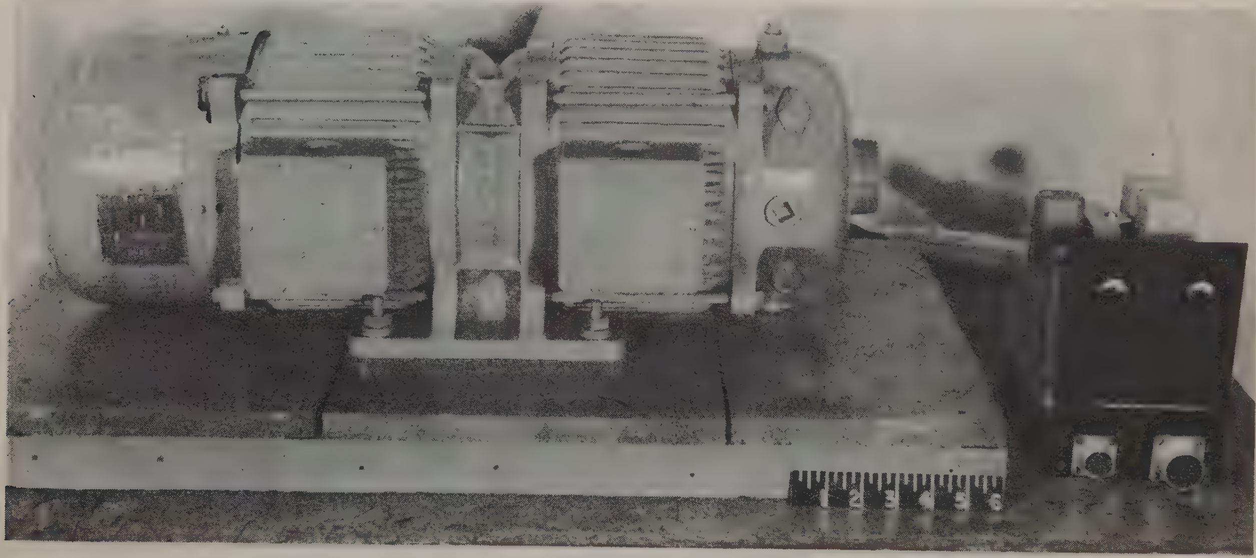


Fig. 9

For higher power ranges, two stages of generation may be necessary in order to get adequate power gain, but it has been demonstrated that this arrangement is preferable as compared to compensating for the time constants involved in large size amplidynes. A sample of such a generator is shown in Fig. 9. A A third approach is to provide a series of one speed motors which can be incorporated into servos to eliminate the gear train. Work covered to date in this area has been discussed in the previous section.

There is one other area in which hope for improved performance lies and that is in the operation of the position servo as a combination velocity and position system. This arrangement supplies a gun order signal which includes both a position signal and a velocity signal. The fundamental scheme to be considered centers around the fact that a velocity system has 90° more permissible phase shift within the closed loop and hence can tolerate more time constants than a position system. Therefore, if our feedback system is built up on the basis of a velocity system and sufficient gain is introduced to maintain dynamic shaft correspondence, there will be little difference observed in the performance of the system except for drift as compared to a position system. The drift problem can then be corrected by means of a low

gain stable position tie. Such equipment has been built and demonstrated in the case of naval directors and experimental work on army tank stabilization units. To design a servo of this type, there seems to be a need for an integral synchro/tachometer combination to simplify the order signal transmission. Problems of linearity and voltage gradient must still be examined before adequate equipment of this nature can be built.

CONCLUSION

A review of the type of performance which can be expected in modern feedback systems indicates that progress will be made by improving components. In this paper much emphasis has been placed upon the activities in the field of naval equipment where horsepower demands are large and where size and weight are not as dominating as in the case of aircraft equipment. We are not overlooking the aircraft problem, however, and in certain areas have already made attempts to develop the hydraulic analogs which will bring lightness of weight to the same applications which are apparent in naval equipment. The development of improved hydraulic components holds much promise that we will be able to do this.

AUTOMATION AS THE ENGINEER SEES IT*

W. R. G. Baker
General Electric Company
Syracuse, New York

Summary -- Problems of automation are discussed with particular emphasis on their importance to society and to the engineer.

What makes automation such a fascinating topic to engineer or social scientist, manager or labor leader, is that each views automation from the perspective of his own experience. This perspective, sometimes creates distortion. However you think of automation depends entirely upon your point of view; to my mind, the answer lies considerably to the right of center.

How much my outlook toward automation is colored by my engineering background is a matter for conjecture. I sometimes find myself agreeing with the economist rather than with the technical man. In a recent Harper's article, economist Peter Drucker spells out a number of truths about automation. First, he states that automation is something that is here now, not something that is approaching; secondly, that automation is not just the pushbutton factory. Mr. Drucker then states that there are three basic principles that make up the logic of automation: economic activity as a process; the pattern, order or form behind changing economic phenomena; and finally, self-regulating control. He closes his article with the statement, "Automation is not gadgeteering, it is not even engineering; it is a concept of the structure and order of economic life, the design of its basic patterns integrated into a harmonious, balanced and organic whole."

In technical terms, and limiting the definition simply to the mechanical aspects, automation is the addition of the tools of programming and computing to the well used tools of mechanization -- mechanics, hydraulics and pneumatics. If this be automation, how then can it also be "a concept of the structure and order of economic life"? I think the apparent disparity in viewpoints can be reconciled. To the engineer, the machine is predictable, because of the physical laws that govern its actions. To the engineer, society is often unpredictable because action and reaction are sometimes not t immediately apparent or may be obscured by the reaction to a combination of forces.

To the economist, society is not unpredictable if all of the economic forces at work can be discovered and identified. And if the self-regulating factors influencing society are recognized, then "feedback" becomes one of the essential elements in the structure.

To the engineer, then, automation relates merely to the machine; but the engineer can see that the utilization of the machine may have far reaching economic effects. This may become a disturbing factor in the engineer's

*This paper was presented at the Centennial Symposium, School of Engineering, Michigan State College, East Lansing, Michigan, on May 13, 1955.

thinking, for he is not certain that what he is building is beneficial for or detrimental to society. It becomes advisable for the engineer to broaden his thinking, so that he may see the "animal" as a whole, and not as a mere machine. Automatic machinery, or the automatic factory or office, when seen in proper prespective, becomes one result of automation, in this broader concept, and not a primary cause of automation.

Automation is not a thing of the future; it is here and is rapidly increasing in scope. We must talk of it in terms of the present.

There are instances where consumer products plants are now turning out appliances at the rate of one every 30 seconds, with 2500 machines strung over 25 miles of conveyors. In data processing, large computers have taken over the work af handling payrolls for thousands of employees. As many as three million accounts of a national insurance company are now serviced by machines using electronic principles. In department stores, new controls supervise as many as 40 remote operations, tying in lighting systems, air-conditioning and a host of other units which previously required human guidance.

The list of instances where automation is already used in one form or another is long. But the surface has hardly been scratched. In terms of the nation's total production, the output through automation is now but a pittance.

Because automation and electronics are inseparable, it will definitely be a dominating factor in our own industry in the era ahead. Five years ago sales of automatic control devices by all companies amounted to a billion dollars. In the past couple years, the figure has tripled, and it is still increasing.

One recent survey showed more than 750 companies are now making control systems, components and equipment. The entrance of so many new firms into the field is predicated on the awareness that eventually all of industry must automate to some greater or lesser degree. There can be no doubt that many companies using self-regulating equipment will obtain a marketing edge over their competitors. The other must follow suit or fall by the wayside. When an employer does not automate, he faces bankruptcy and his workers face the loss of their jobs.

In industry automatic production is based on the idea of unbroken motion and involves the combination of several piecemeal operations into one unit. By using the "building block" technique, you link a number of automatic systems, gradually broadening the process until the completely automatic plant is within view.

The ultimate, of course, would be the pushbutton type of operation covering inspection, assembly, test, packaging and a raft of other services in which no human help is needed. We know of no completely automatic factory operating on a large-scale basis at the present time. The closest approach to it is the processing of some fluids -- in the chemical, oil and food industries -- where only elaborate piping, pumping and time cycling is required. One leading organization of chemical producers has estimated that instruments and control mechanisms now commercially available can supply 90% of the needs

of the modern chemical plant. This is illustrative of the great things possible with automation. But let us remember that while these things can be done, it doesn't mean they will be done.

Complex machines to make improved goods quicker are costly. In the hurried pace of the future, marketing habits will change and some machines are bound to become obsolete. Management must look at the situation from the standpoint of the number of hours a specific machine can do the required job before it has to be altered. There was an instance of this in the automobile industry last fall. You may be familiar with the story of the car manufacturer who went all-out for automation in producing engines -- to a certain point. Workmen still inserted semifinished connecting rods by hand into a device carrying them through transfer machines. Here's a case where available automatic equipment could have solved the particular problem. But at $90,000 it cost too much, and the manufacturer resorted to a combination of automatic and manual principles to get his finished product.

On the other hand, while looking at the expense factor, there's the case of the company in the East which recently contracted for an electronic brain at a monthly cost of $30,000 -- $360,000 a year -- and figures it will be money well spent. In a matter of minutes, the machine supplies information that formerly required hundreds of manhours. The process used to take so long that figures were obsolete by the time they were computed. Now the system is used to speed the whole operation to the point where extra clerks are needed to keep business rolling once the desired answer is obtained.

We continue to hear rumblings from people who feel that the machines of automation will bring contagious unemployment. They've compared it to a highly communicable virus that will have all of the nation's workers flat on their backs, when, in reality, the minor readjustments accompanying the early stages of automation should be considered as being sort of a vaccine. Early immunization with the mild dose will bring lasting protection in the long run. In automation, we can look at the minor temporary dislocations as being but a step toward the stabilization of employment in the future.

We have heard all kinds of suggestions from those who claim automation has its darker side. Using the theory that when machines replace men, the men who have no jobs cease to be customers and throw other men out of work, some people have even suggested that the government stop issuing patents on automatic devices. If such a step had been taken 15 or 20 years ago, when mechanization was growing by leaps and bounds, the workers now employed in the field of electronics might have been deprived of their present jobs. At GE alone, there are 45,000 people today who are making products that were unheard of 15 years ago. Their jobs exist only because we have employed the principles of mechanization and automation as a means of lowering consumer costs.

In face of this, we still have some people who insist automation is a sinister menace leading to economic catastrophe through disintegration of the labor market. Fortunately, however, the realization is now beginning to take hold in some quarters that the monster is nothing more than a Halloween spook -- once you take off the false face, the person underneath isn't bad at all. To those who have fostered automation as a step to a better way of life, there is no surprise in recent pronouncements by some resistance leaders that

standards of living can be raised considerably if the problem is handled correctly.

Take 1930 as a base year for the past quarter of a century of progress. Since then, production of machinery has increased about four times. Our gross national product has done the same. Retail prices have doubled, but at the same time, average industrial earnings for employees have more than tripled. The number of hours worked each week has gone down and the number of people working in industry is now double the figure of 25 years ago. The portion of those doing hard, unskilled work has been cut in half. Better incomes have given the workers more purchasing power. Shorter hours have meant more pay for less work, and more leisure time.

The whole concept of our new economy is that cheaper prices across the counter will enable the customer to satisfy basic needs, creating additional purchasing power. Past experience has shown that, as needs decrease, wants intensify. Knowing consumers will require more goods than we can now produce, our developmental laboratories must begin to anticipate the markets of the future.

The potential is great. In the next seven years, the population of the United States may go as high as 184 million. The increase -- 24 million -- gives us a tremendous market to cultivate.

Despite the pessimistic outlook of some everything we have read in recent months points to a need for an augmented labor force in the years ahead, not a smaller one. Due to the low birth-rate of the 1930's the number of people entering the work force will be slim -- some of the experts have estimated it at only 11%. This, at a time when we must double production if we are to meet the requisites of the growing market. If we do not use automatic machinery, how else are we to achieve this goal?

From this, you might conclude that we have nothing but clear sailing ahead. To convey this impression would be to present a picture that is not entirely true. You can produce as much as you want with automation, but eventually the customer will get a yen for a different product. If he doesn't, it's up to you to provide the necessary stimulation. Otherwise, your market becomes saturated and you come to a dead end.

The very fact that we must resort to new designs to maintain an even market should, in itself, serve as a handbrake against the quick emergence of the wholly automatic plant. In applying the advanced technology, the manufacturer must measure the pace of progress to determine how extensive the application should be.

The key to widespread use of automation is flexibility. The ability to switch quickly from one type of assembly to another, with a minimum of downtime on equipment, is a prime requirement. This is especially true in the case of the small supplier, who is pressured into a race to keep up with the big producer's demand for new parts. We at GE have come a long way in this respect with our automatic component assembly system which is adaptable to the needs of the big producer and small manufacturer alike.

That's one of the faults we can find today with almost every discussion taking place on the subject of automation generally. Too many people believe automation's broadest use will be in the field where only the big boys play. But the "little fellow" is definitely in the picture. In some cases, his problems will be more intense than those arising from the progressive process of the larger plants. In some instances, however, he will be spared many of the headaches. For example, he will not have to tie up large amounts of existing capital in research and creative engineering. That's a proposition the larger manufacturers will encounter and will have to measure up to as the use of automation widens.

To achieve solid commercial success, those companies using new techniques in great degree must eventually look to the public for new money. If we are to entice the consumer to part with what he has saved through automation's s cost reductions, we must give them new designs to even out the purchase pattern. The sound conduct of business demands it. This is particularly true if fast-moving fields like electronics, where all the knowledge at our command must be sharpened to bring forth new designs to satisfy the appetite of the public for better living.

This responsibility is squarely on the shoulders of industry's engineers. Already, hundreds of ideas are coming from the drawing boards. There's one inherent danger, however. We must not only satisfy the whims of today's buyer -- the fellow whose tastes we know -- but of the potential purchaser of five to ten years from now. The public will dictate whether an item is obsolete before it gets to the production stage. With our new economy, we could well be forced into a situation compelling more frequent introduction of new items. It means our engineers will have to pay closer attention to deadlines and performance dates in the future. Those who have been disrespectful of deadlines in the past may find there is no longer any time for head-scratching. In getting the originality we seek quicker, there may be an extension of the present trend to "team engineering."

With the strain for fresh ideas, we must assume that there will be a need for more engineers than we now have. For that reason, we may be rightfully disturbed by some of the statements made recently by those who have pinned the responsibility on management to provide the means of moving displaced production people to new jobs. While reaching for a greater share of the profits, they make no allowance whatsoever for the sizeable amounts of capital which must be diverted to engineering research to create the products which will make those new jobs available. Some industries now have a ratio of as many as 100 engineers for every 1000 workers. There is every reason to believe that, with automation, this proportion will be changed considerably.

While much of our emphasis has been on the laboratory engineer, we must be mindful, too, of the fact that the manufacturing engineer will take on somewhat of a different appearance as advancements are made. As automation comes in, there will be less of a need for specialization and a greater requirement for those trained in a number of engineering lines. We will require controls specialists to supervise entire assemblages of apparatus, so the complicated machines will remain in proper working order. They will have to know more than the intricate details of instruments and electronics and will be required, too, to know a good deal about general machine operations -- like welding, stamping, and other functions.

Much of this knowledge will be acquired on the job. But, in a greater measure, it must be supplied by our schools of technology. Systems engineering has opened a whole new field of technical education.

I think the facts speak for themselves. We need not fear that which has so entranced the science fiction writers for some decades, "the invasion of the robots." We do have a need to recognize that the acceleration of technology, which marches hand in hand with the concept of automation, places the engineer even more firmly on the management team. It places greater responsibilities upon his shoulders, for he must, as a member of that management team, help determine whether the product must be designed for automatic production. The decision cannot be made on the basis of "Can it be done?" but "When should it be done?" with full consideration for the human, the marketing, the financial and the profit problems involved.

The responsibility also rests upon the engineer to help find a solution to the great and growing need for increased technical education and for finding the incentives that will attract qualified young people into technical fields.

As a member of the management team, the engineer must help see to it that we do not become so wrapped up in the magic of the word "automation" that we ignore some of the areas that can provide equally important rewards to our economy and our standard of living. Management must find better methods of handling marketing and distribution and human relations problems.

IRE ACTIVITIES IN THE FIELD OF AUTOMATIC CONTROL

(As mentioned in the editorial, the IRE has contributed to the development of feedback control for a number of years, but only recently has it officially entered the automatic control field. This official activity, particularly in establishing control standards, is described briefly by J. E. Ward, present chairman of IRE Subcommittee 26 on Feedback Control Systems. -- Editor's Note)

IRE Technical Committee 26 on Feedback Control Systems was established in 1951 under the chairmanship of Professor William M. Pease of the Massachusetts Institute of Technology. Among the initial objectives of the Committee were the establishment of standards on letter symbols and terminology for feedback control systems, methods of measurement and the promotion of an IRE professional group in this field. Most of these initial objectives have been attained. The IRE Professional Group on Automatic Control was formed in 1954 under the chairmanship of Robert B. Wilcox, an original member of Technical Committee 26. The first two standards prepared by Committee 26 were approved in 1955 by the IRE Standards Committee. The first of these, "Graphical and Letter Symbols for Feedback Control Systems -- 1955," appeared in the November, 1955, issue of Proceedings of the IRE. The second standard, "Terminology for Feedback Control Systems," appeared in the January, 1956, Proceedings.

STANDARDS AND THE CONTROL FIELD

In the IRE organization, the technical committees, working under the Standards Committee, are responsible for the preparation of all standards of terminology and methods of measurement. A committee cannot publish its work until completed and approved by the Standards Committee and the IRE Executive Committee, nor can it sponsor meetings. On the other hand, a professional group cannot work on standardization, although it can sponsor meetings and suggest areas or material for consideration to the technical committees. The committees and the groups thus complement each other.

As one example of the need for standardization in the control field, consider the name of the committee itself. Initially it was called the Technical Committee on Servo-Systems. Later, the name was changed to Feedback Control Systems, because this had become a more widely used term for the field. The parallel AIEE committee had made this same name change earlier. Whence comes the name "automatic control" used by the new professional group? This is an even broader term for our field and is also used as the name for ASME and American Standards Association committees in this field. This difficulty in even naming the field of interest indicates the need for adequate standards.

The original committee consisted of six members, all of whom are still active. The current membership is seventeen. The Committee has at the present time one Subcommittee 26.1 on Terminology for Feedback Control Systems, under the chairmanship of M. R. Aaron of the Bell Telephone Laboratories which has been working since 1952 on terminology. Mr. J. C. Lozier was the organizer and the first chairman of this subcommittee. A second subcommittee on methods of measurements is in the process of being formed.

The current committee membership is:

J. E. Ward, Chairman   E. A. Sabin, Vice-Chairman

M. R. Aaron	D. L. Lippitt
G. S. Axelby	J. C. Lozier
G. A. Biernson	T. Kemp Maples
R. E. Graham	W. M. Pease
V. B. Haas	P. Travers
R. J. Kochenburger	R. B. Wilcox
D. P. Lindorff	S. B. Williams
W. K. Linvill	F. R. Zatlin

The current membership on Subcommittee 26.1 is:

M. R. Aaron, Chairman

G. S. Axelby	T. Flynn
G. R. Arthur	J. C. Lozier
V. Azgapetian	C. F. Rehberg

F. Zweig

With the completion of the initial standards on symbols and terminology, the Committee is proceeding with the preparation of additional standards and material on standards of performance and methods of measurement. The Committee feels that the expanding interest of the IRE in automatic control warrants intensive effort in these standardization activities. In addition to their activities on Committee 26, a number of committee members are also active in similar committees of other professional societies and in the American Standards Association. This direct coordination of the work of committees of the various societies will hasten the realization of universal standards in the control field.

John E. Ward
Chairman, Committee 26.0

On October 5, 1954, the IRE Board of Directors approved the petition for the formation of a Professional Group on Automatic Control. Sponsorship of this professional group was by members of the IRE Technical Committee on Feedback Control Systems and members of a number of IRE sections in answer to desires for a specialized organization which would be active in all phases of automatic control. A definite need was felt for an organization that would effectively aid in the tremendous progress being made in the field of automatic control by making contributions at the technical sessions of the IRE conventions, by organizing local professional chapter meetings, by publishing papers in the Transactions on Automatic Control, and by sponsoring national technical symposia.

The purpose of the PGAC was given considerable thought in choosing the field of interest which appears in the constitution:

> "The field of interest of the Group shall be automatic control systems. It shall encompass the components thereof, such as transducers, data transmission links, computers and control devices and the integration of these components into control systems. It shall include scientific, technical, industrial or other activities that contribute to this field, or utilize the techniques or products of this field, subject, as the art develops, to additions, subtractions or other modifications directed or approved by the IRE Committee on Professional Groups."

The main aspect of the Group's activity is the integration of computers, data processing, electronic techniques, feedback control, switching circuits, telemetering, etc., all under the broad umbrella of automatic control. Such things as inventory control, manufacturing, accounting, and even business operations, can be considered control situations in that they are amenable to automatic operations. Yet the scope has been maintained broad enough to cover the more general control problem with which the feedback control engineer is familiar. The real challenge is the integration of component parts and subsystems into a complete control system.

There are three main areas made available to the PGAC members (at present close to 1200) where they can obtain and disseminate information on the theory and practice of design of control systems: the national group symposia and technical sessions on a national or regional scale, the local professional chapters and the IRE Transactions on Automatic Control.

To date the PGAC has sponsored technical sessions at a number of conventions. At the 1955 National Convention the program included a technical session on automatic control and the panel session entitled "Trends in Automatization of Procedures and Processes in Business and Industry," with Dr. G. S. Brown of the Massachusetts Institute of Technology as chairman. The national group and the Dallas-Fort Worth Chapter sponsored the technical session on automatic control at the South Western IRE Conference and Electronics show in February, 1955. Again in August the national group with the Los Angeles Chapter sponsored the automatic control session at Wescon. These activities will become an annual event and will be supplemented by additional convention programs. When our Group becomes sufficiently strong, we plan to

initiate a symposium on automatic control, possibly in conjunction with one of our sister societies.

Within the professional group organizational structure are the local professional chapters. The chapters have their own organization with a chairman, vice-chairman, secretary-treasurer and activities committee, but maintain close coordination with the local sections. Such an organization has the definite advantage of providing local activity in one's professional field by means of meetings, panel sessions, etc., which are sponsored by the chapter or with other chapters in closely allied fields of interest. In this way, the members can take an active part in the group functions. A dozen or more active chapters will provide a strong national organization which in turn will be able to provide more for its members. During 1955, chapters have been organized in Baltimore, Boston, Dallas-Fort Worth, Los Angeles and Twin Cities. At present the New York Chapter is being organized as well as the Philadelphia, Connecticut Valley, San Diego and San Francisco. Monthly meetings have been arranged by these chapters to provide technical sessions for their members. Most of the papers presented to local organizations are also made available for publication in the Transactions on Automatic Control or the IRE Proceedings.

The 1955 Convention Record Part IV contained those papers presented at the convention and represented the first published material of the PGAC. This spring 1956 issue of the Transactions on Automatic Control is our first Group publication. It is hoped that two or three such issues will be brought to the members during 1956 and on a quarterly basis thereafter. The success of the PGAC and its publication depends on the members to supply the Papers Procurement Committee with technical material.

I would like to take this opportunity to inject a note of appreciation to the Administrative Committee, the Standing Committee members and Chapter Officers for their cooperation and effort since the formation of PGAC. A concerted effort on the part of all members during the coming year will be necessary to increase our technical meeting activity and to provide quarterly issues of the Transactions.

Robert B. Wilcox
Chairman, PGAC

INFORMATION FOR AUTHORS

George S. Axelby
Westinghouse Electric Corporation
Baltimore, Maryland

Summary--Because the first PGAC Transactions papers will be reproduced from photographs of original manuscripts, it will be necessary to maintain a standard form to provide uniformity, neatness and legibility.

This manuscript is written and typed in the desired form. Detailed instructions for authors are included.

INTRODUCTION

Until the PGAC has grown sufficiently large, it will be necessary to publish the Group Transactions in the most inexpensive manner. This will be done by reproducing photographs of original, typewritten manuscripts and black-ink line drawings. To achieve a desirable uniformity and legibility, the manuscripts should be typed with the same margins, spacing and headings wherever possible.

Ordinarily a paper should contain the following sections: "Summary," "Introduction," main section (with appropriate headings), "Discussion," "Appendix" and "References." These will be considered individually in more detail.

ARRANGEMENT OF MATERIAL

Summary

An abstract of no more than 200 words precedes the article and is called "Summary." This section informs the reader briefly of the subject, method and results of the work. The summary should be informative and should outline the essentials of the paper. References or equations are not included. The essence of a good abstract is a concise statement of what was done and why.

Introduction

The purpose of the introduction is to orient the reader with respect to the problem. This section may be especially brief in a Transactions paper as much of the historical and introductory material will generally be known to Transactions subscribers, specialists in the field.

Body of the Report

In the main section or body of the report each step in the development of the material should be described as clearly and briefly as possible. Long mathematical developments are to be placed in an appendix -- not in this section. Only a mathematical outline should be given and its importance and role in the development of the work clearly stated.

Of course this section should not be headed "Body of the Report." But although it need not follow any particular form, for clarity there should be appropriate informative headings (and subheadings if necessary).

Discussion

After the work has been described and the results have been given, the significance of the work is discussed. In this part of the report some of the conclusions stated in the summary are developed. This section of the paper is important. Failure to state conclusions clearly is a common fault and the significance of a worthwhile contribution may be lost because the conclusions have not been clearly presented to the reader.

Appendixes

The appendixes include detailed subdevelopments of the main discussion.

References

A complete list of references should be given at the end of the paper and proper notation with superscripts made throughout the text. Each reference should be complete. For an article in a periodical the information contained should be: author's name, title of article, name of periodical, volume number, page numbers, month and year of publication. For a book: author's name, title of book, city of publication, publisher, year of publication, page numbers.

Illustrations and Tables

Tables, photographs and diagrams are to be arranged so that they are self-explanatory and can be used with a minimum of reference to the text. The quantities and units used in plotted curves should be given clearly and the legends worded to convey as much information as possible to the reader who may be casually glancing through the *Transactions*.

MARGINS, MEASUREMENTS AND STYLE

All manuscripts are to be typed single space on anpelite typewriter (twelve characters to the inch) on $8\frac{1}{2}$ by 11 inch paper. The width of the text should be 6 3/4 inches exactly (81 characters) and the depth 8 3/4 inches.

Titles and Subtitles

The title of the paper should be concise, but clear. The author's name and affiliation are given after the title. The city and state of the author's affiliation are also included, but street addresses and zone numbers omitted. No degrees or titles are mentioned.

All main headings are also centered and capitalized, except for "summary." This particular heading is typed upper and lower case, underscored, and flush with the left-hand margin. A dash follows and the text begins on the same line. (See first page of this article.)

Subheadings (or side headings) occupy a separate line, are underscored and upper and lower case. A line is skipped before and after the subheading.

Illustrations and Tables

All tables must be typed and not drawn. Diagrams must be original drawings in black ink on white paper or tracing cloth; photographs must be glossy prints.

Their size should not exceed $8\frac{1}{2}$ by 11 inches. Drawings with typewriting on them are not acceptable. All information must be hand-lettered in black ink, the lettering large enough to stand reduction. In graphs only the major coordinate lines should be drawn; a chart containing a large number of closely spaced lines will not reproduce well.

Footnotes

All footnotes are placed at the bottom of the page on which they are indicated. The first footnote is designated with an asterisk (*), the second with a dagger (†) and the third with a double dagger (‡).

SUBMITTING PAPERS

The review and publication of papers will be seriously delayed if manuscripts and illustrations are not submitted in proper form. The original and two copies should be sent to:

Mr. M. R. Aaron
Bell Telephone Laboratories
Murray Hill, New Jersey

Papers submitted should include a statement as to whether the material has been copyrighted, previously published or accepted for publication elsewhere. Material not acceptable for publication will be returned.

DISCUSSION

Using the information here and following the instructions as closely as possible will result in neat, legible PGAC Transactions that should convey valuable information to the control engineer in a minimum amount of time.

COMMITTEES AND CHAPTERS OF THE IRE PROFESSIONAL GROUP ON AUTOMATIC CONTROL

PGAC Committees

Membership Committee..................	V. Azgapetian -- Chairman
Paper Study and Procurement Committee............................	M. R. Aaron -- Chairman G. A. Biernson W. K. Linvill P. Travers
Publication Committee.................	G. S. Axelby -- Chairman and *Transactions* Editor V. Azgapetian -- 1956 Convention Record
Chapter Activities Committee..........	R. B. Wilcox
1956 Convention Program Committee.....	J. C. Lozier -- Chairman
Awards Committee......................	G. A. Biernson -- Chairman

PGAC Chapters

Baltimore.............................	G. S. Axelby -- Acting Chairman
Boston................................	P. Travers -- Chairman G. A. Biernson -- Vice-Chairman T. Slattery -- Secretary-Treasurer
Dallas-Fort Worth.....................	A. R. Teasdale -- Chairman C. D. Buehrle -- Vice-Chairman W. L. Evans -- Secretary-Treasurer
Los Angeles...........................	E. M. Grabbe -- Chairman J. Salzer -- Vice-Chairman K. Fish -- Secretary-Treasurer
Twin Cities...........................	P. M. LaHue -- Chairman W. S. Wells -- Vice-Chairman G. J. Murphy -- Secretary-Treasurer
New York..............................	(Being Organized)